

# **Indo-European, Nostratic, and Beyond: Festschrift for Vitalij V. Shevoroshkin**

---

Editors

*Irén Hegedűs, Peter A. Michalove  
and Alexis Manaster Ramer*



# **Indo-European, Nostratic, and Beyond: Festschrift for Vitalij V. Shevoroshkin**

---

Editors

*Irén Hegedűs, Peter A. Michalove  
and Alexis Manaster Ramer*

**Journal of Indo-European Studies**  
Monograph Number 22  
**Institute for the Study of Man**  
1133 13th St. N.W. #C-2  
Washington D.C. 20005

**Indo-European, Nostratic, and Beyond;  
Festschrift for Vitalij V. Shevoroshkin**

**Editors:**

*Irén Hegedüs, Peter A. Michalove  
and Alexis Manaster Ramer*

ISBN 0-941694-59-3

**Journal of Indo-European Studies Monograph Series**

*Editors: A. Richard Diebold, Jr. and Edgar C. Polomé*

**Number 22**

Copyright 1997

Institute for the Study of Man Inc.

1133 13th St. N.W. #C-2

Washington D.C. 20005

*Tel: (202) 371-2700 Fax: (202) 371-1523*



## Table of Contents

Introduction	iii
Vitalij Viktorovich Shevoroshkin: Selected Publications	iv
Raimo Anttila: Beating a Goddess out of the Bush?	1
Václav Blažek: Indo-European 'Seven'	9
Claude Pierre Boisson: The Phonotactics of Sumerian	30
J. C. Catford: The Myth of the Primordial Click	51
Madhav M. Deshpande: Pāṇini and the Distinctive Features	72
Joseph H. Greenberg: Does Altaic Exist?	88
Eric P. Hamp: A Far-Out Equation	94
Irén Hegedűs: On Grammaticalization in Nostratic	106
Pramila Hemrajani: Three Kisses	116
Peter Edwin Hook: Relative Clauses in Eastern Shina	140
Vyacheslav Vs. Ivanov: Luwian Collective and Non-Collective Neutral Nouns in <i>-ar</i>	155
Brian D. Joseph: Macrorelationships and Microrelationships and their Relationship	168
Mark Kaiser: Rigor or Vigor: Whither Distant Linguistic Comparison?	183

Leonid Kulikov: Vedic <i>mriyáte</i> and other pseudo-passives: Notes on an Accent Shift	198
Alexis Manaster Ramer: The Polygenesis of Western Yiddish -- and the Monogenesis of Yiddish	206
Karl Heinrich Menges: Etymological Problems with Words for 'Blood' in Nostratic and Beyond	233
Peter A. Michalove: Altaic Evidence for Clusters in Nostratic	243
Vladimir Orel: New Albanian Etymologies (Balkan Etymologies 116-145)	257
Ilya Peiros: Macro Families: Can a Mistake Be Detected?	265
Richard A. Rhodes: On Pronominal Systems	293
Merritt Ruhlen: Proto-Amerind *KAPA 'Finger, Hand' and Its Origin in the Old World	320
Sergei A. Starostin: On the "Consonant Splits" In Japanese	326
Alexander Vovin: Some Japanese Etymologies	342

## Introduction

The career of Vitalij Viktorovich Shevoroshkin has spanned two continents, and some four decades. Perhaps this explains the remarkable breadth of his work, and the variety of topics addressed by the contributors to this volume. Beginning his training in the Soviet Union in the 1950s, Vitalij became a specialist in Anatolian linguistics, and made important early contributions to the study of the recently discovered Carian inscriptions. His work on Anatolian lead to the study of other languages of the middle east, particularly Afroasiatic.

His emigration from the Soviet Union in 1974 led him through a series of European universities until he reached his present home at the University of Michigan in 1977. There, he has continued work in a variety of fields, including Nostratic and other proposals for long-range comparison. But aside from his own work in this field, he is best known to many of us for his efforts as an advocate of these exciting new ideas in historical linguistics. In this, he has always served as a champion of his former compatriots, Aron Dolgopolsky and the late Vladislav M. Illič-Svityč.

The possibilities of long-range linguistic comparison are still in their early stages, and many of the ideas that have been proposed so far undoubtedly need significant revision, while some may ultimately be discarded altogether. But the study needed to separate the wheat from the chaff can take place only if these ideas are heard and discussed, and our honoree has probably done more than anyone else to promote the discussion and dissemination of a wide variety of proposals. His efforts have brought the field of long-range comparison from the days when such proposals were received with little more than indifference, to the present state, where evidence for a variety of ideas is vigorously presented, debated, and gradually refined. Whatever the final judgement on many of these ideas, we are all indebted to him for what we have learned from them.

More recently, Shevoroshkin has returned to his roots in more ways than one. Just as the end of the cold war has facilitated continued contacts with his Russian colleagues, so he has recently returned to his early interest in Carian and Anatolian studies.

So, Vitalij Viktorovich, whatever your next endeavor, we wish you continued success and offer you this volume as thanks in this, the year of your 65th birthday.

Peter A. Michalove, Irén Hegedűs, and Alexis Manaster Ramer.  
January 1997

## Vitalij Viktorovich Shevoroshkin

### Selected Publications

1957

“K istorii indoevropskogo genitiva.” *Voprosy jazykoznanija* 6:89-90.

1958

“K Probleme rekonstrukcii.” *Filologičeskie nauki* 3: 56-61.

1962

“Karijskij vopros.” *Voprosy jazykoznanija* 5:93-100.

1963

“O strukture zvukovyx cepej.” In *Problemy strukturnoj lingvistiki*: 164-181. Moscow.

1964

“Karijskij jazyk: Sovremennoe sostojanie dešifrovki i izučenija.” in *Problemy indoevropskogo jazyko-znaniya*: 18-39. Moscow.

“Karijskij: odin iz xettskix jazykov xetto-lidijskoj podgruppy.” in *Problemy sravnitel'noj grammatiki indoevropskix jazykov*: 62-84. Moscow.

“Novoe issledovanija po xettologii.” *Voprosy jazykoznanija* 3:124-133.

“On Carrian.” *Revue hittite et asianique*: 1-55.

“Zur karischen Schrift und Sprache.” *Kadmos*: 72-87.

1965

*Issledovanija po dešifrovke karijskix nadpisej*. Moscow.

“Maloazjajskie jazykovye paralleli.” *Ètimologija*: 142-59.

“O xetto-lidijskom xarakteru karijskogo jazyka.” *Vestnik MGU* 5:45-57.

1966

“K Probleme sootnošenija zvukovyx i fonemnyx cepej.” in *Issledovanija po fonologii*: 160-165. Moscow.

1967

*Lidijskij jazyk*. Moscow.

1968

"Zur hethitish-luwische Lexik." *Orbis* XVII, 2:467-91.

"K Probleme likijskogo jazyka." *Voprosy jazykoznanija* 6:66-80.

"Zur Entstehung und Entwicklung der Kleinasiatischen Buchstabenschriften." *Kadmos* VII, 2:150-173.

"Karisch und Lykisch." in *I. Congress di mecenologia*, I:462-472. Rome.

1969

[with A. Koroljov]. Lykische Wörter und Namen." *Archiv orientální* 37:523-542.

*Zvukovy cepi v jazykax mira*. Moscow.

"Zu den späthethitischen Sprachen." *Zeitschrift der deutschen Morgenländischen Gesellschaft*, supp. v. 1: 250-271.

"Zur Erforschung der kleinasiatischen Onomastik." in *10. Internationaler Kongress für Namenforschung*: 341-350. Vienna.

1970

[with A. Kondratov]. *Kogda molčat pis'mena*. Moscow.

1971

"Zur Erforschung des Milyischen." *Mitteilungen des Instituts für Orientforschung* XVII, 2:206-36.

1974

[with A. Kondratov]. *Zagadki velikoj sem'ji*. Moscow.

1975

"Zur sidetischen Schrift und Sprache." *Kadmos*: 154-167.

1976

[with A. Koroljov]. "Xetto-luvijskie jazyki." in *Jazyki zarubežnoi Azii*: 13-93. Moscow.

1977

“Zu einigen karischen Wörtern.” *Münchener Studien zur Sprachwissenschaft* 36:117-130.

“Zu einigen Verwandtschaftsbezeichnungen im Lykischen und Milyischen.” *Münchener Studien zur Sprachwissenschaft* 36:131-140.

1978

“Sull’ interpretazione delle righe 20-21 della trilingue de Xanthos.” *Incontri linguistici* 4/2:238-239.

“Studies in Hittite-Luwian Names.” *Names* 26, 3:231-257.

1979

“On the Hittite-Luwian Numerals.” *Journal of Indo-European Studies*: 177-98.

1981

“On the Hittite-Luwian and IE Etymologies.” in *Bono Homini Donum*, ed. Y. Arbeitman, Part 1:259-62. Amsterdam: John Benjamins.

1982

“Zu den hethitisch-luwischen Konsonanten.” in *Investigationes philologicae et comparativae: Festschrift für H. Kronasser*: 210-214. Wiesbaden.

1984

“Theophoric Names in Carian.” *Onomata* 9:9-16.

“Über den Lautwert des karischen Buchstaben Q.” *Incontri linguistici* 8:71-78.

“Verbesserte Lesung von karischen Wörtern.” *Incontri linguistici* 9:199-200.

1985

[with Mark Kaiser]. “On the Indo-European Laryngeals and Vowels.” *Journal of Indo-European Studies*, v. 13 nos. 3/4:377-413.

1986

[ed. with Thomas Markey]. *Typology, Relationship and Time: A Collection of Papers on Language Change and Relationship by Soviet Linguists*. Ann Arbor: Karoma.

[with Mark Kaiser]. “Inheritance versus Borrowing in Indo-European, Kartvelian, and Semitic.” *Journal of Indo-European Studies*, v. 14 nos. 3/4:365-78.

1987

[with Mark Kaiser] "Indo-European and Afroasiatic." *General Linguistics* 1:34-46.

[with Mark Kaiser] "On Recent Comparisons between Language Families: The Case of Indo-European and Afro-Asiatic." *General Linguistics*, v. 27 no. 1:34-36.

Indo-European Homelands and Migrations." *Folia linguistica historica*, v. 7, no. 2:227-250.

1988

"On Laryngeals." in *Die Laryngalthorie*, ed. A. Bammesberger: 527-546. Heidelberg: C. Winter.

"Indo-European Consonants in Anatolian." in *A Linguistic Happening in Memory of B. Schwartz*: 283-303. Louvain-la-Neuve.

[with G. Barinova] "Some Observations on the Role of *oslyshki* in Linguistic Research." *Russian Language Journal*, v. XLII, nos. 141-143:309-329.

[with Mark Kaiser] "Nostratic." *Annual Review in Anthropology*, v. 17:309-329.

"Carian Proper Names." *Onomata* 12:497-505.

1989

[ed.] *Reconstructing Languages and Cultures*. Bochum: N. Brockmeyer.

[ed.] *Explorations in Language Macrofamilies*. Bochum: N. Brockmeyer.

"Methods in Interphyletic Comparisons." *Ural-Altaische Jahrbücher*, v. 61:1-26.

1990

[ed.] *Proto-Languages and Proto-Cultures*. Bochum: N. Brockmeyer.

"Identities and Semi-Identities in Deep Reconstruction." in *Evolution: Molecules to Culture*. New York: Cold Spring.

[With Alexis Manaster Ramer] "Some Recent Work on Remote Relations of Languages." in *Sprung from Some Common Source*, Ed. Sydney Lamb and Douglas Mitchell: 178-199. Stanford: Stanford University Press.

1991

[ed.] *Dene-Sino-Caucasian Languages*. Bochum: N. Brockmeyer.

"On Carian Language and Writing." in *Perspectives on Indo-European Language, Culture and Religion: Festschrift for Edgar Polomé*, v. 1: 117-135. Washington: McLean.

[with J. Woodford]. "Where Linguistics, Archaeology, and Biology Meet." in *Ways of Knowing*: 173-197. New York: Prentice Hall.

#### 1992

[ed.] *Nostratic, Dene-Caucasian, Austric and Amerind*. Bochum: N. Brockmeyer.

#### 1993

"Lykisch und andere spätanatolische Sprachen." in *Akten des zweiten internationalen Lykien-Symposiums*: 38-52. Vienna: Österreichische Akademie der Wissenschaften.

"Carian: 30 Years Later." in *The First International Conference on Carian*. Rome: University.

"Identical Words in Anatolian Languages." in *The Second International Congress of Hittitologists*. Pavia: University.

#### To appear

[ed.] *Studies in Nostratic*. Bochum: N. Brockmeyer.



## Beating a Goddess out of the Bush?

Raimo Anttila

*University of California, Los Angeles*

Suppletion is normally an indication of great historical depth and of basic vocabulary. I have suggested such a situation for the roots we know as *\*ag-* 'drive' and *\*g<sup>w</sup>hen-* 'beat (down upon), kill' (Anttila 1986): On the one hand the former root signals hunting, fishing, seizing, and even killing in quite a number of contexts, and on the other, Balto-Slavic and Albanian lose it and assign also the general driving meaning ('drive') to the latter, if *\*weġh-* does not do the work. But best of all, the two roots are in actual suppletion in Hittite where *ak(k)-/ek(k)-*, curiously only active in form, acts as the semantic passive to *kuen-/kun-* 'kill', and means rather 'to be put/sentenced to death' (cf. *φόνος* 'death as punishment' [Soph.], and also Skt *han* means 'put to death, cause to be executed, punish'), which is of course easily epiphenomenally 'die', and this gets emphasis in the handbooks. I will not repeat here the quite interesting Hittite details, nor the Balto-Slavic rich gamut in which 'driving' divides nicely into 'herding' and 'hunting-chasing-beating'. Both roots hark back to (paleo)lithic times, i.e., to hunting and gathering, where both aspects fall under beating, whether battue-beating or throwing together nuts and berries. When such an economy shifts to agriculture, old terms can be carried over, and normally would be carried over. Thus the Slavic *e*-grade *\*žen-* 'to reap' has been lexicalized into an independent root, but it nicely reflects non-hunting beating, whereas *gonobit*, (*s*)*gonošit* 'gather, save' would refer to preservation of any goods acquired (*žatva* 'zapas - storage', in addition to 'crops, grain', and even 'gain, profit'; *žniva* 'stubble-field, crops'; cf. Mägiste's idea that Finnish *aitta* 'granary' < *\*ajitta ~ aja-* 'drive'). Parallels are easy to find, cf. Swedish *slå hö* 'to mow' (beat hay) and Finnish *tappaa riittä* 'thresh' (beat the *riihi* [threshing barn]).<sup>1</sup> Again, we find parallel reflexes on the *\*ag*-side, e.g. Gothic *akran* 'fruit, crop' (*\*agro-no-*

---

<sup>1</sup>Today *tappaa* is in all other contexts 'kill', and provides thus a perfect parallel to *\*g<sup>w</sup>hen-*. Note further Vedic *han* with *áva* and *prāti* as 'thresh'.

*m*; cf. ἀγείρω 'gather' and ἀγρέω 'take, seize', e.g. in trapping). Similar fruit terms are attested in Celtic. Of course, the standard opinion is that such terms do not belong here, but I find their relevance quite attractive, and as I will try to show below, perhaps also productive. Such a situation is taken as high-level proof of explanation in historical inferences (such productiveness takes the place of prediction in natural science), and linguistic change is no different. It is also part of history, and explanation therein is inherently historical.

As with the decendants of PIE *aĥ-* 'sharp', which bear witness to stone-age technology, we might have a similar situation in *\*g<sup>w</sup>hen-* pointing to stone-age economy. In a hunting and gathering situation, abundance and riches and life itself is food, what you are able to beat together. Fick did make the proposal (approved by Bechtel) that ἄφενος 'riches, abundance' would essentially be *\*sm-g<sup>w</sup>henos*, going nicely with εὐθενέω 'thrive, flourish' (~ εὐθηνέω). It seems that the general idea ever since has been that the root meaning here would be 'swell', rather than *\*'beating together'*, duly considered by Szemerényi (1964:144-6; I let this serve as the basic locus for references) who reminds us that φόνος αἷματος 'blood clot' must belong here (see fn. 5). But he finds an *-es-* stem compound noun *\*sm-g<sup>w</sup>hen-es-* a "wholly artificial construct" unacceptable for phonetic development as well. I suppose that the phonetic problem is the labial for the expected dental, but this is not that big a difficulty.

Let us go farther afield and look at Homeric ἄφαρ 'suddenly, quickly', usually (and rightly) connected with ἄφνω 'of a sudden'. This looks like perfectly good heteroclisis with *\*-ɾ-/-n-*, and there has been a good parallel in εὐθύς 'straight, immediately, at once' and εἴθαρ 'at once, forthwith' with a parallel *\*-ɾ-/-u-*, with the suggestion that the former has assimilated the original front glide diphthong to the following back vowel. I think it is rather the reverse, i.e. dissimilation of *eu* before another labial, as in *\*we-wk<sup>w</sup>-o-m* > εἶπον 'I said', although admittedly there is more lip-rounding in the latter case; in both cases, however, the *u* is unstressed, which seems appropriate.<sup>2</sup> In my mind the adjective is in fact a cousin of εὐθετος 'quick, able' (Demosthenes). Note a fair

<sup>2</sup>The problem of *ei* > *i* is a separate one (e.g. ἱθύς) and has no bearing on the issue here.

and accurate parallelism in *\*dhE-to-* ~ *\*-dhE-u-* and *\*plE-no-* ~ *\*plE-u-*. Now we might get a compound without the *-es-* by analyzing ἄφαρ as containing *\*sm-g<sup>w</sup>h<sub>2</sub>h-* \*'[with] one blow'; cf. German *plötzlich*, originally 'auf einem Schlag', and earlier also *slage slags* 'mit einem Schlag, plötzlich, auf einmal'. As in πυκνός/πύκα we would of course expect *\*ἄφα*, which would indeed be a good adverbial shape (μάλα, ἄμα, τάχα, etc.). Instead of staying with ἄμα 'at the same time', *\*ἄφα* would now have taken its ending from εἶθαρ resulting in ἄφαρ. Lexicographers list even a further accrual in ἄφαρ εἶ, and indeed, such remodeling is quite commonplace in this semantic nook. Ἄφνω now looks like an original instrumental, and it might even reflect an original athematic formation; very old features are often fossilized in adverbs.

The verbs εὐθενέω and εὐθηνέω are obviously denominative, but it is not certain that they are from the *-es-* stem *\*-θεν-εσ-*, and in any case there is a different prefix. This might in fact be quite significant. As I said, particularly in a stone-age conception, life and riches are one and the same thing. Koivulehto (1991:36-44) draws our attention to this again by treating harvest words from the root shape *\*os-*, *o*-grade of 'to be', *\*es-*. He also points out the Finnish parallel of *elo* 'harvest; goods, property' to *elää* 'to live'. Goods is also the outcome of the other 'be' root, *\*wesu* > Skt *vasu* (cf. *was/Wesen*). Koivulehto further presents Finnish and Permian developments of PIE *\*Eesu-* (Skt *ásu*, Greek *ἔύς*) in *\*kese* 'gut, tüchtig, passend; Freund, Gatte' (Lat. *erus* 'master', OLat. *esa* 'mistress'). When we now take this noun as a possible component of the compound, i.e. *\*E(e)su-g<sup>w</sup>hen-*, we might indeed have one of the original contexts of the Greek prefix. This would be something like \*'beating out the sustenance (= life)', putting it together, in other words, 'abundance and riches'.<sup>3</sup> Tautology of this kind is a strong indication of the original meaning, it is parallel to compounds like *lemon-yellow*. Ἄφενος < *\*sm-g<sup>w</sup>hen-* would share the same semantic field as *\*E(e)su-g<sup>w</sup>hen-*. We would actually not know when the *-es-* stem was formed, because it would have been easily possible after the compound had faded, or the boundary was

<sup>3</sup>In fact, φερέσβιος 'life-bearing, life-giving, nourishing [earth]' is a good portrait of this kind of semantics. And in connection with *erus* it would be nice to be able to prove that εὐπάτωρ (and εὐπολις) derives from the same situation. Note the (paradoxical historical) contrast between *\*Aoyu* 'life' > οὐ 'not' and *Eesu* 'life' > εὖ- 'good'!

blurred. As for the compound, this is exactly what we saw in Russian (*s*)*gonošit*’, even with or without *\*s(o)m-*, in exactly the right home economy reading. Sanskrit *sam+han*, in addition to the regular killing and destruction readings, means something like *sam+dhā*, joining, putting together, beating together, making compact. Note particularly *saṃhati* ‘keeping together, saving, economy; bulk, heap, multitude’ (close to *-fassa*; see below). If we now could trust the Greek, we could satisfy Meillet’s three-witness requirement in syntactic reconstruction, it seems. But at least the semantic field can be strengthened from Vedic, through *vrj* ‘twist off, pluck, break somebody’s neck’. This is the root apparently cognate with German *werfen* ‘throw’, and note that *han* covers such a meaning with *ā* (on which more below), *ud*, and *ni*. But particularly in the context of sacrificial grass *vrj* means ‘gather’, and generally ‘choose for oneself, select’, and *sam+vrj* ‘lay hold of, seize for oneself, appropriate, own’ (which comes quite close to ἄφρεος), i.e. to throw booty together, and the nominal forms echo this: *saṃvargá* ‘rapacious, gathering for oneself’ (~ *saṃvārgam*, *saṃvārjana*). Collection of sacrificial grass as part of religion could go back tens of thousands of years, although it is difficult to prove, of course.<sup>4</sup>

The swelling meaning is there also, in other words, it is a natural outcome of driving/beating.<sup>5</sup> Grassmann’s dictionary lists *sam+han*

<sup>4</sup>The importance of the original Proto-(and Pre-)Indo-European nature religion also comes out well in Haudry (1987). This treats the Hera cluster and gives strong background and support for the Demeter/Persephone aspects below. Nature and plants in the original hunting and gathering culture shimmer also in Greek sports (Sansone 1988).

<sup>5</sup>This might not be quite clear by just juxtaposing English *ache* vs. Swedish *åka* ‘drive’, but the parallels from Finnish (e.g. *ajos*) and Karelian make it an obvious possibility (see Anttila 1986). It is also quite dubious to keep a root ‘swell’ and one for ‘beat’ separate (both *\*g<sup>w</sup>hen-*). Thus Skt *āhanás* ‘swelling’ can quite well be (\*)‘heranschlagend’ (cf. German *Ausschlag*), with *ā* ‘near (to), toward’ not that far from *\*som-*. Φόνος αἵματος supports the semantics delineated, but particularly Russian *vygon(ka)* ‘distillation, burning tar, driving to the pasture’ (i.e. *āhanás* in the meaning ‘pressing out [soma]’), and the verb *vygonjat* means also ‘destroy’ (note further *vygn[a]ivanie* ‘festering, rotting’).

With *abhi+ā+han* we get ‘beat, kill’. We would actually like more information on *āhanás*, but it is clear that its meaning is something like ‘lascivious’, and the term refers to copulation (and from this one gets the *āhanasyās*, obscene

as a milking term (which would fit the milk and honey metaphor), but the passage goes (RV 8.31.9c) *sám ūdho romaśam hato* 'they press together the udder and the hairy one', thereby doing their duty to the gods. Here we have metaphors for the female and the male genitals, and these could have been formed any time. The result (of "beating [it] together") is of course children who in their time uphold the prosperity of the community and secure worshippers for the gods, etc., but there is no unambiguous original prosperity meaning here. On the other hand, good being/life is working together, good union. And good union can be many things indeed. Koivulehto quotes (from the \*Eesu side) from Finnish dialects *kesy* 'wer sich allzu leicht mit dem anderen Geschlecht vertraulich macht' (1991:40) and *kesu flikka* 'ein Mädchen, das den Jungen willig ist' (43) (*Geschlecht* is cognate with *schlagen* ~ *slay*!).<sup>8</sup>

\*ΑΦΕΝΟΣ agrees with the hunting and gathering starting point in that it reflects the cattle-raising and agrarian counterparts (as do εὐθενέω and εὐθηνέω). In Homeric the meaning is tied to grain and cattle, i.e. plants and animals as concrete riches rather than abstract richness. The adjective ἀφνειός refers to individuals and their houses, not cities, which seems to indicate that originally it was a good beater that was "rich" (and his possessions were kept in his house)<sup>6</sup>, and that beating it together for the common good was on a different level.<sup>7</sup> Practically all words in the nature-gain-crops domain can develop into profit/prosperity aspects, so it is very important to try to see the earliest connotations. The handbooks

"copulation verses"). How much swelling is necessary cannot be determined, since the reading could also be metaphoric from *ā+han* 'to stick (the axle) in (the wheel), to beat/pound violently'. But now the problem is that Yama accuses his sister of the property, and this does not seem to fit the axle fitting, because scholars have looked at it only from the axle side. But the action is identical when looked at from the wheel perspective. Note that in Modern English *horny* refers also to women, and in Black English women have a *cock*. In the light of Baltic Finnic parallels it is possible that *ajá* 'he-goat' was in fact \*'fucker'. Note also Russian *gon* 'rut' (cf. *Trieb*) and sexual meaning for the verb also.

<sup>6</sup>Πολυφόντης would be such an individual, but names are indeterminate. Even if in \*Αργειφόντης we have killing in the second part, it need not be true in the starting point of the former. \*ΑΦΕΝΟΣ goes into names in Thessaly, e.g. Τιμαφένης (Szemerényi, p. 144).

<sup>7</sup>Curiously, the adjective is the epithet of Ares in Arcadia, in the meaning of 'the nurturing one', almost like Skt *bhara* (and with passive meaning: *bhāryā*, *bharita*), but then also 'booty, battle' (cf. fn. 9).

today keep -φεν- and -θεν-/θην- apart, but there is no good reason for it. Of course the latter look the same as Latin *fēcundus* - *fēlix* - *fēmina* - *fētus*, and these are taken under *\*dhē-*. In between fall further Latin *fēnus/-oris* 'interest, gain, profit' and *fēnum* 'hay', again with an ambiguous -n- (going with the root or the suffix?). Fick analyzed the latter as *\*fend-snom* 'abgemähtes' (cf. *dē-fend-ere* 'beat off'), and in this context it becomes again quite attractive.

Formally an -es-stem is not that unique after all, if one considers Latin *Venus*, in which the stem type remains even after it has been personified as a goddess with female grammatical gender. Here we have another parallel to the material under discussion, since *\*wen-* is perhaps an original hunting term with some ties to plants.<sup>8</sup>

Murder, slaughter, and blood are meanings (of e.g. φόνος) that easily result from the battue-beating context, or the hunting aspect. The problem is, and has been, the gathering (or the later agricultural) aspect which has left only vague remnants. If we assume that the action meaning shifts to the result of action we will get a rather natural solution for the long-standing problem of the name of Persephone, on which enough has been written (and I will not review it here in any form whatsoever). There are quite a number of forms, e.g. Περσεφόνηα, -φόνη, -φασσα/-φαττα, not to speak of the problems of the first part (with an aspirated initial). And there is no dearth of suggestions for etymology. But if we take, say, *\*g<sup>w</sup>honā* and *\*g<sup>w</sup>hntyā* (> Oīc. *gunnr* 'battle' [ultimately the source of *gun*]) as nature's abundance or some such, then the first part works nicely in connection with πέρθω 'waste, ravage, sack' (as has indeed been suggested), i.e. the name means basically the disappearance of the earth's riches or abundance or the profitable gain to be had, which fits in nicely with the time she spends downstairs (= loss of grain and game).<sup>9</sup> There is general agreement

<sup>8</sup>The root *\*wen-* gives an incredible rich gamut of meanings of joy and lust (and cf. again Hittite *wenzi* 'fucks', and German *Wonne*), profit (*Gewinn*, *Gewinst*), but note particularly Gothic *winja* 'pastur(age), fodder' and ON compounds with *vin* 'meadow' or some such (e.g. *Vinland*). The semantic distance to Swedish *vän* 'friend' seems considerable.

<sup>9</sup>As I said, many basic terms end up with produce/gain/crop meanings, which is natural in a stone-age economy. Fraenkel's analysis took the gain component in reverse order. He found an -s-aorist of φέρω in φερσε-, which is embarrassing in that such a form is not otherwise attested. Like Hesychius' ἡ φέρουσα τὸ ἄφρονος it would be 'pregnant with riches' (but with the swelling part at the

that the Demeter/Persephone complex contains much from the pre-Greek culture, but such a situation need not mean that the name itself could not in essence be inherited. And would the violation of Grassmann's Law in the dialectally frequent first part  $\Phi\epsilon\rho\sigma\epsilon-$ / $\Phi\epsilon\rho\rho\epsilon-$  hark back to a phrasal compound, or to a time when there were phrases like those suggested here to counteract the deaspiration law (cf. the imperative with  $-\theta\iota$ )? Such an early semantic cohesion could also explain the generalized labial in  $\tilde{\alpha}\Phi\epsilon\nu\omicron\varsigma$  (which is hardly weirder than the variants (with "flitting" aspiration)  $\Phi\epsilon\tau\tau\alpha\lambda\omicron\varsigma$  and  $\Pi\epsilon\tau\theta\alpha\lambda\omicron\varsigma$  for 'Thessalian').

It is of course true that basic terms go into new metaphors and contexts, cf. *to throw up a log cabin*, Swedish *slå ihop* 'put [= hit, beat] together (e.g. income)', or Finnish *lyö-dä leiv-i-lle* (beat-INF. bread-PL.-ALLAT.) 'will do', etc., and thus all such features need not be inherited. But it seems that in this case (of hitting nature for food to religion) the most obvious semantic possibilities have been ignored. Our honoree has shown in the Nostratic context that strict observance of semantic change and sound laws can certainly reach tens of thousands of years back. In this vein I believe we can indeed reap more from Greek. In beating the bush in such bold and successful footsteps I have been able to scare up just a few possibilities, but I hope they do him honor nonetheless.<sup>10</sup>

## References

Anttila, Raimo

- 1986 "Deepened joys of etymology, grade A (and Ä)". *Journal de la Société Finno-Ougrienne* 80: 15-27.

Haudry, Jean

- 1987 *La religion cosmique des Indo-européens*. Milano and Paris: Archè.  
[Earlier versions as: "Héra - 1", *Etudes Indo-Européennes* no. 6.17-46 (1983) — "Héra - 2", *Ibid.* 7.1-28 (1984). — "Héra et les

---

end). Indeed,  $\Phi\omicron\rho\alpha$  is also 'produce, crop, fruit' (cf. *Ertrag*). It would seem that this was originally what nature carried and put forth, and  $\tilde{\alpha}\Phi\epsilon\nu\omicron\varsigma$  what man was able to throw together ( $*\Phi\omicron\nu\alpha$ ). In the name of the goddess this distinction is gone, as would be expected in such a complex religious tradition behind it.

<sup>10</sup>As so often before, I thank Hanns-Peter Schmidt for invaluable advice on things Vedic.

héros", *Ibid.* 12.1-5 (1985). — "Les Heures", *Ibid.* 18.1-14 (1986).]

Koivulehto, Jorma

1991 *Uralische Evidenz für die Laryngaltheorie*. Vienna: Austrian Academy of Sciences.

Sansone, David

1988 *Greek athletics and the genesis of Greek sports*. Berkeley, etc.: University of California Press.

Szemerényi, Oswald

1964 *Syncopé in Greek and Indo-European and the nature of Indo-European accent*. Naples: Istituto Universitario Orientale.



# Indo-European "Seven"

Václav Blažek  
Příbram, Czech Republic

Dedicated to Professor Shevoroshkin, in the middle of the seventh decade of his life.

1. The numeral "7" is well attested in all branches of Indo-European:

Indo-Iranian:

\**septm̥* "7" > Old Indic *saptá*, Pali *satta* (cf. "Mitanni-Aryan" *šatta* in Kikkuli's text), Hindi etc. *sa:t*; Kati *sut*, Waigali *so:t*, Ashkun *su:t*, Prasun *sētē*, Khovar *sot*, Kashmiri *sath* etc.; Avestan *hapta*, Khotanese *hauda*, *hoda*, Pashto *o:wa*, Sogdian *β t( )* = \**avd*, Yaghnobi *avd*, *aft*, Alanic *αβδα* [in *Ἀρδαβδα*, lit. "(city) of seven gods", the proper name of the city of Theodosia], Ossetic *avd*, Yidgha *ávdo*, Shugni *(w)u:vd*, Wakhi *hü:b* etc., Parachi *hö:t*, Zoroastrian Pahlavi, Modern Persian *haft*, Kurdish (Kurmanji) *hävî*, Baluchi *apt* etc.

\**septm̥-mó-* "7th" > Old Indic (AV, YV and exclusively in classical Sanskrit) *saptamā-*; Khotanese *haudama-*, Khwarezmian *βdym*, Sogdian *βtm(yk)* = \**avdami:k* (cf. personal names *Ἀφθαίμακος*, *Ἀφθειμακος* known from Tanais), Ossetic Iron *ävdam*, Parthian *hftwm*, Zoroastrian Pahlavi *haftom*, Modern Persian *haftum*.

\**septm̥-t(H)o-* "7th" > Old Indic (only RV) *saptátha-*; Avestan *haptaθa-*. Emmerick (1992a: 182) sees in *saptátha-* the secondary form based on the reinterpretation of *ṣaṣthá-* "6th" as cardinal plus suffix *-thá-*. Elsewhere he differentiates the Indo-Iranian suffixes *\*-tha-* : *\*-ta-*, interpreting them as the specific opposed to the general respectively (1992b: 323). Schmidt (1992: 198) takes account of the identity of the suffix of Old Indic ordinals "4", "5", "6", "7" and the superlative, assuming their common pronominal origin.

\**septm̥-tf-* "70" (orig. "Siebenheit"; cf. Debrunner and Wackernagel 1930: 369, 419; Mayrhofer 1996: 681 for *ṣaṣṭi-* "60") or \**septm̥-(d)k̑ntH₂* > \**sapta(:)ćati-* > Old Indic *saptatí-*; Avestan *hapta:iti-* (but *haptaiθiuuant-* "seventyfold"), Khotanese *hauda:tä*, Manichean Sogdian *βt't*, Khwarezmian *βd'c*, Pashto *awia:*, Ormuri *awaitu*, Middle Persian (Turfan) *hpt'd*, Zoroastrian Pahlavi, Modern Persian *hafta:d* etc. (Abaev 1958: 82-83, 196-197; Bailey 1979: 498-499; Berger 1986: 29; Emmerick 1992: 169-170, 175, 181-182; Id. 1992b: 299, 310, 323; Mayrhofer 1976: 431; Id. 1996: 700; Morgenstierne 1927: 13).

## Anatolian:

\**septmiyo-* or \**septm̥-yo-* > Hittite *siptamiya-* "a liquid consisting of seven ingredients", cf. 3-*ya-al-la* 7-*mi-ya* *šipantanzi* doubtless corresponding to *ši-ip-ta-mi-ya te-ri-ya-al-la šipandanzi* "sie libieren *siptamiya* und *triyalla*", i.e. liquids consisting of seven and three ingredients resp. (Kronasser 1966: 169, 365). Eichner (1992: 85) explains the change \**e* > *i* by *i*-umlaut. He finds a formal parallel concerning \*-(*i*)yo- extension in the Roman name *Septimius*. The form *siptamiya-* is a derivative of an original ordinal \**siptama-* < \**septmó-* (Eichner 1992: 84; let us mention an alternative reconstruction, \**septm̥mo-*). The unextended *o*-stem is probably preserved in the Cappadocian female name *ša-áp-ta-ma--ni-ga*, which has been interpreted as "the seventh sister". The *a*-vocalism indicates most likely a Luwian source, cf. Luwian *sap(pa)tammimalli-* "sevenfold" (?), interpreted as the participle of an unattested denominative verb *sa(pa)tammiya-* "to multiply by seven" (Melchert 1993: 188). Shevoroshkin (1979: 190) tries to add Milyan *sejtamiu*, attributive to *qirzē* (acc. sg.) "share", identifying it on the basis of other attributes *tbiplē* "double" and *trpplē* "triple" with Hittite *siptamiya-*. The irregular change \**pt-* > *-jt-* can be explained by the influence of *aitāta* "8".

\**septm̥(t-)* > \*[*se/ipt*]an- > Hittite 7-an "7" (Eichner 1992: 83-84).

## Armenian:

\**septm̥* "7" > Armenian *ewt'n*. In the variant *eo:t'n* < \**ewt'n* the contamination of *ewt'n* and the dialect form \**awt'n* may be viewed (Winter 1992c: 350). Kortlandt (1994: 254) prefers to see here "...a reduced grade vowel, which replaced zero grade vocalism in the ordinal and was later introduced into the cardinal."

\**septm̥-(d)kontH2* "70" > \**ewt'an-sun* > \**ewt'asun* > Armenian *ewt'anasun*. Winter (1992b: 352-353) assumes that *-n-* was introduced from "7" and the cluster \**-wt'n-* was reduced in complexity by the insertion of *-a-* before *-n-*. Kortlandt (1994: 255) sees in *-asun* (also *k'ar:asun* "40") the phonetic reflex of \**dkont-* (he reconstructs \**dkomt-*) after a syllabic resonant.

## Greek:

\**septm̥* "7" > Greek *heptá*.

\**septmo-* "7th" > \**sebdmo-* > Ionian-Attic *hébdomos* (with *-o-* inserted under the influence of *ógdo(w)os* "8th" ?), Delphian, Cyrenaean, Aetolian *hébdemos* (*-e-* is puzzling; see Waanders 1992: 380). Szemerényi (1960: 8, 12, 93) reconstructs a different development: \**septm̥mos* > \**heptamos* > \**hebdamos* (with *-bd-* after "70") > *hébdomos* (with *-o-* after "8"). The

Homeric alternative form *hebdómatos* perhaps follows *tétratos* (beside *tétartos*) "4th" < \**k<sup>w</sup>etr-to-* similarly as *trítatos* "3rd".

\**septm-dkontH<sub>2</sub>* "70" > \**septmH<sub>1</sub>kontH<sub>2</sub>* > \**hebdmé:konta* > Greek *hebdomé:konta*, Delphian, Heracleian *hebdemé:konta* (Waanders 1992: 375, following Kortlandt 1983: 98-99; Beekes 1995: 214 accepts the originality of -a < \*-H<sub>2</sub>, contrary to Kortlandt and Waanders). Sommer (1951: 23) judges that -é:- was introduced through "60" from "50". Kortlandt l.c., starting from the glottal theory, explains -é:- in *penté:konta* "50" by compensatory lengthening as follows: \**penk<sup>w</sup>e-dkont-* > \**penk<sup>w</sup>e-'kont-* > \**penk<sup>w</sup>e-H<sub>1</sub>kont-* > \**penk<sup>w</sup>eekont-* > \**penk<sup>w</sup>é:kont-* (cf. also Waanders l.c.).

Dacian:

\**septm* > Dacian \**sipta* and -a:k(o)s > \**siptoax* > *sipotax* and *sipoax* (Pseudoapuleius) "heptápleuron, septenervia; Wegerich" (Georgiev 1977: 196-197; as a formal parallel in word formation he quotes Bulgarian *sedmák* "seven years old animal").

Albanian:

\**septm-ti-* > \**septá-ti-* > \**se(p)tá-ta:* (for the replacement of \*-ti- suffix forming numeral abstracts by \*-ta: > -të - see Hamp 1992: 912) > \*š(ě)tá-të (the form *šët-* is preserved in Lakonia and Triphylia Arvanitika in *e šëtúnë* "Saturday", normally *e shtunë*. See Hamp 1992: 894) > Albanian *shtatë* "7" (Hamp 1992: 914). Mann (1977: V) finds in the Illyrian (?) proper name *Stataria* a possible reflex of pre-Albanian numeral "7".

Italic:

\**septm* "7" > Latin *septem*.

\**septm-mo-* "7th" > Latin *septimus*, earlier *septumo* (CIL 1.2519); cf. Marsian proper name *Setmiu[s, Setm]ius* = Latin *Septimius*.

\**septm-dkñteH<sub>2</sub>* "70" > \**septmH<sub>1</sub>kñteH<sub>2</sub>* > \**septma:genta:* > \**septuma:ginta:* > Latin *septua:ginta:*. (Coleman 1992: 395-396, 401-402, 411-412).

Celtic:

\**septm* "7" > Insular \**sextem* > Old Irish *secht N*; Brythonic \**sextam* (with irregular \*-s- instead of expected \*-h-) > Middle Welsh *seith*, Cornish *seyth*, *syth*, Breton *seiz*.

\**septm-eto-* "7th" > Gaulish (La Graufesenque) *sextametos* (< \**sextam* + \*-etos after *pinpetos* "5th"), Middle Welsh *seithvet*, Cornish *seythves*, Breton *seizved*; Old Irish *sechtmad*.

\**septm̥mo-(d)konts* "70" > Old Irish *sechtmogo*. (Thurneysen 1946: 250; Lewis and Pedersen 1954: 235, 239; Bernardo Stempel 1984: 140; Greene 1992: 510, 515, 540).

Germanic:

\**septm̥t* "7" (with *-t* after the ordinal \**septm̥to* - ?) > \**sepmt̥* > Germanic \**sebun* > Gothic *sibun*, Crimean Gothic *sevene*; Old High German *sibun*, Old Saxon *sibun*, *sivon*, Old Frisian *sigun*, *sōgun*, *sowen* etc., Old English *seofon*(n), *seofun*, *siofu*(n), *sifu* etc. ; Old Icelandic *sjau*, Old Swedish *sju*., Danish *syv* etc.; the preservation of *-t* in *septun* (Lex Salica) has been explained by Latin influence. Hamp (1952: 138) assumes the following development: \**septm̥* : \**septm̥to* -> early Germanic \**seftu* : \**sibun*ðaz and after leveling of cardinal on analogy to ordinal \**sibun* : \**sibun*ðaz. Szemerényi (1960: 35) proposes an original solution explaining the loss of \**-t* based on metathesis \**seftun* -> \**sefunt* -.

\**septm̥to* - "7th" > Germanic \**sebun*ða- > Old High German *sibun*, Old Saxon *sibondo*, Old Frisian *sigunda*, Old English *seofopa*; Old Icelandic *sjaundi*, Old Swedish *siundi* etc.

\**septm̥-dé:k̥nt*- or -*dék̥nt*- "70" > Gothic *sibuntehund* (Ross and Berns 1992: 609). Among other explanations (cf. Lehmann 1986: 301; Shields 1992) the solution of Szemerényi (1960: 33-35) is doubtless the most sagacious: \**septm̥:kont*- > \**seftun*χanþ- > \**seftune:hund* (after the operation of Lex Verner and influenced by \**χunþan* "100") > \**seftune:hund* (after \**fimfe:-hund* "50") > \**sefunte:hund*.

\**septm̥-dék̥nt* "70" > Germanic \**sebun-tegu*- > Old Saxon *sibuntig*, Old High German *sibunzug*, *sibinzig* etc. Old Icelandic *siau tiger*, Old Danish *siutiugh*, Old Swedish *siutighi* etc. (Ross and Berns 1992: 602-609, 617). The distinctive reconstructions \**dék̥nt* and \**dék̥nt* are justified elsewhere. The other, more complicated forms (Ross and Berns 1992: 618) are not important for our purpose to study the numeral "seven".

Baltic:

\**septm̥* "7" > Baltic \**septin* + *-i*: (after \**keturi*: > *keturì* "4") > East Baltic \**septi:ni*: > Lithuanian *septynì*, Latvian *septiņi*, dial. *septīņi* (Smoczyński 1989: 84; concerning \**-i*: > *-i* he quotes Old Lithuanian *patì* "wife, female" < \**pati*:, cf. Old Indic *pātni*: "lady", pp. 98-99, fn. 15). Stang (1966: 279) explains the lengthening of the second vowel by analogy of *aštuoni* "8" < \**ašto:-*. Yatwingian *gepti*š "7", correctly probably \**ʃepti*š (Zinkevičius 1984: 12), can reflect \**septins*.

\**septm̥-mo-* "7th" > Balto-Slavic \**septima-* (or \**septuma-*) > Baltic \**septma-* > Old Prussian *septmas* (II, III 1x), f. *septmai* (III, 1x), *sepmas* (III, 1x); East Baltic \**setmas* > Old Lithuanian *ŝėkmas* (the substitution \*-tm- > -km- can be illustrated e.g. by *šálkmėtės* "mentha piperita" < \*šált-métēs or by *áukmonas* "boss" < German *Hauptmann* per Smoczyński 1989: 84), *Sekminės* "Whit, Whitsunday" (Fraenkel 1962-65: 772).

\**septm̥-to-* "7th" > East Baltic \**septin-ta-* (after \**devin-ta-* "9th") > Lithuanian *septiñtas* (an innovation appearing only in the end of 18th cent.), Latvian *septītais*.

Slavic:

\**septm̥-mo-* "7th" > Balto-Slavic \**septima-* > \**septma-* > pre-Slavic \**sebdmũ* > West and South Slavic \**sedmъ* and East Slavic \**semъ*. The cardinal \**sedmъ* originated after the ordinal \**sedmъ* replaced the expected but unattested \*\**setь* or \*\**setę* a regular continuant of Balto-Slavic cardinal \**septin* (Lamprecht 1987: 121-122). Comrie (1992: 756-757) offers an alternative solution consisting in coalescence of cardinal \**setь* < \**septin* < \**septm̥* and ordinal \**semъ* < \**septmo-*, giving \**setmъ* > \**sedmъ*. The unique Kashubian forms *sětäm*, *sětma* with voiceless -t- have been explained as a result of regular devoicing before -m̥ (Comrie 1992: 756).

Tocharian:

\**septm̥* "7" > \*šəpət(əN-) > A šäpt(äN-) > pl. šäptäntu, in compounds šäpta-, after metathesis špäť; B šäwät > \*šwät > \*šut > šukt after *okt* "8" (Winter 1992b: 109). Van Windekens (1976: 461) presents a traditional solution for the B form: \**septm̥* > \* šäptäm̥ > \*šäptu > \*šäktu (after \**aktu* > *okt* "8") > \*šukt.

\**septm̥-to-* "7th" > \*šəpətəNtV > A šäptänt, B šuktante and šuktänte (Winter 1992b: 137-138; he notices a formal identity of Lithuanian *septiñtas*).

\**septm̥-(d)kñtH2* "70" > \*šəpətəNka > A šüptuk (with -u- after *oktuk* "80"), B šuktanka (Winter 1992b: 121).

## 2. Reconstruction and etymology.

The preceding analysis confirms the traditional reconstruction of the indeclinable cardinal \**séptm̥* (Beekes 1995: 215; the accent shift in Aryan-Greek-Albanian-Germanic \**septm̥* reconstructed by Brugmann 1892: 478 was probably

caused under the influence of the numeral "8"; see Debrunner and Wackernagel 1930: 356 with older literature; Schwyzler 1939: 590) and ordinal *\*septm-mó-* (more probable than *\*septmó-*). Other reconstructions do not respect the facts, e.g. *\*sepŋt* is acceptable only for Germanic (Voyles 1987: 492; cf. also Shields 1992: 89, 97), and in the case of *\*sequdm* < *\*seque* "apart" and *\*duo*: "2" Mann (1984-87: 1129-1130), assumes the change  $k^w > p$  not only for *p*-Celtic, Osco-Umbrian and post-Mycenaean Greek, but for all Indo-European branches.

In spite of the tempting possibility to identify the final *\*-m* with accusative, the consonant stem *\*sept-* ("heptad" ?) or only the root *\*sep-* remain etymologically unanalyzable (Winter 1992a: 12); the attempt of Schmid 1989: 13-14 to see here the *\*-ti-* derivation from the root *\*sep-* with the original meaning *"Pferde mit Hand und Zügel zusammenhalten"* cannot be accepted for semantic reasons; similarly unconvincing are the attempts of his predecessors as Pott, F. Müller, Stewart (see Debrunner and Wackernagel 1930: 356). Studying the systems of numerals in various language families, I am convinced that it is possible almost always to determine an original motivation of all higher numerals beginning "5". For the case of missing etymology the following rule can be formulated: If a numeral *x* in a language *A* has no hopeful etymology and there is a similar numeral *x'* in a neighboring language *B* where *x'* is analyzable, the question of the borrowing  $x < x'$  is quite legitimate. It is remarkable that the numeral "7" in most of the language families in the neighborhood of Indo-European resemble the form *\*septm* studied in 1.

### 3. External parallels

#### A. Uralic languages

a) Fenno-Permic *\*šeŋćemä* (Joki 1973: 313; Rédei 1988: 773), *\*še(e)ś/ćVmi* (Sammallahti 1988: 553), *\*šejććem* (Honti 1993: 100-102; he admits also *\*s-*), *\*šeć(ć)em(ə)* > *\*še:ćem(ə)* > *\*šejćem(ə)* (Napolskikh 1995: 126); Balto-Fennic *\*sejććen*, *\*sejććemä-* (after Honti 1993: 102); Finnish and Ingrian *seitsemän*, dial. *seitsen*, Carelian *seittšemen*, *seittšimä*, *seittšen*, Olonets *seit't'še(i)*, Weps *seittšmen*, *seičmen*, Wote *seitse:*, gen. *seitsme:*, Estonian *seitse*, gen. *seitsme* etc.

Lappic *\*će:ćem* > Inari *čiččam*, Norwegian *čiežâ*, Notozero *čihčem* etc. (Lehtiranta 1989: 24).

Mordvin *\*šisəm* (Keresztes 1986: 143).

Merian *\*šeŽum* / *\*šiŽum* (Tkačenko 1989: 121).

Mari *\*šišəm* (Bereczki 1992: 61-62).

Permic *\*šižim* (Lytkin and Guljaev 1970: 255).

The numeral has no hopeful internal etymology. In agreement with the rule formulated in 2. it is natural to seek a source outside Fenno-Permic languages. Among the Indo-European branches contacting Fenno-Permic languages there are two candidates (considered as early as Serebrennikov 1963: 221):

i) Baltic: Old Lithuanian *sėkmas* "7th" allows us to speculate about the source of the type *\*sekma-* > *\*sek'ema-* > *\*šėć(ĉ)em*. The hypothesis of East Baltic origin can be supported by existence of Baltic hydronymy on the vast area between Baltic sea and Volga and by presence of Baltic borrowings not only in Fenno-Volgaic languages but also in Permic branch (Gordeev 1985: 113f).

ii) Slavic: Tkačenko (1989: 121) and Napolskikh (1995: 125-126) see the origin of the Fenno-Permic numeral "7" in Slavic, but it is evident that the hypothetical source cannot be East Slavic *\*semь*. It should be a form very close to *\*setmь* discussed above, perhaps better with the fill-vowel *\*setьmь* (cf. Comrie 1992: 757), which would have had to be transformed into *\*\*šet'čimĩ* (Napolskikh l.c.). The closest parallels within Slavic could be Kashubian *sētām*, *sētma* and possibly Polabian ordinal *sīdim*. The earliest contact of Slavs and Fenno-Permians indicated by archeology is dated to the end of the 4th cent. A.D. (Sedov 1994: 8). A direct connection of these first Slavic immigrants in the North with the basin of middle Vistula is also known (Sedov 1994: 10; cf. Zaliznjak 1988: 176 concerning the linguistic evidence). The main problem remains in chronology. The end of the 4th cent. A.D. is too late for any influence on the common Fenno-Permic proto-language. Sammallahti (1988: 520) puts it between the disintegration of Fenno-Ugric proto-language (3500-3000 B.C.) and the introduction of the Battle Axe culture at 2500-2000 B.C. The only solution would be an independent influence of early Slavic dialect(s) on Fenno-Permic branches, including the possibility of mutual borrowings among them.

The hypothesis of Ross (1941: 1), reconstructing the borrowed Indo-European archetype in the form *\*s/šēk̂sm̂*, a mixture of the numerals "6" and "7", should be also taken in account.

b) Ugric *\*säptä* or *\*sä:ptä* (Joki 1973: 313), *\*θäptə* (Rédei 1988: 844; Honti 1993: 103), *\*Säpt* (Napolskikh 1995: 124; the symbol *\*S* is used for incompatible *\*s/š* > proto-Khanty *\*ǵ* and Hungarian *Ø* and *\*ś* > proto-Mansi *\*s*)

Ob-Ugric *\*θääpet* (Sammallahti 1988: 504), *\*θä:pət* (Honti 1982: 138); Mansi *\*sä:tə* (Honti 1982: 138); *\*s-* < Fenno-Ugric *\*ś-*. The corresponding sound to Khanty *\*ʌ-* is Mansi *\*t-*.

Khanty \**äpət* (Honti 1982: 138); \**ʌ-* < (Ob-) Ugric \**θ-* < Fenno-Ugric \**s-* and \**š-*.

Hungarian \**ét* > *hét* with *h-* after *hat* "6".

Traditionally a donor language has been sought in Iranian (Korenchy 1972: 70; Joki 1973: 313 with lit.). But Iranian \**hapta* could be a source only for Hungarian. The protoform \**θäptə*, common for Khanty and Hungarian with \**š-* < \**s-* (or \**š-*) apparently better resembles Indo-Aryan / Indo-Iranian \**sapta* (cf. Abaev 1981: 85, 89, who rejects the speculations about "early Iranian", yet preceding the typical Iranian change \**s* > \**h*). There are more borrowings esp. in Ob-Ugric, bearing typical Indo-Aryan features, e.g. Mansi LM *šäšwé*, T *ši:šé'ŋ-* "hare" vs. Old Indic *śasā-*, Phalura *šäši:ak* etc., but Avestan \**saŋha-*, Khotanese *saha-* etc. "id." (Blažek 1990a: 42). The expected cultural contact can be localized in time and space: the bearers of the cultural complex Andronovo, very probably speakers of an early Indo-Aryan ("Sauma-Aryans" after Parpola 1994: 156) or even an Indo-Iranian (Kuz'mina 1994) dialect, and proto-Ugrians were neighbors in the contact area of southern Siberia during the 2nd mill. B.C. But the Indo-Aryan hypothesis does not explain Mansi *s-* < \**š-*.

For the vacillation between \**θ* < \**s-/š-* and \**s-* < \**š-* within Ugric an alternative solution can be found in the hypothesis of a Tocharian origin (cf. Joki 1973: 313 "...Zur Klärung des letzteren [= Mansi *s* < \**š*] kann doch. /*šäptä-*/ wohl nicht herangezogen werden: doch. A *šäptänt-* "siebenter"; Janhunen 1983: 120 "...an early Proto-Iranian source is normally assumed [for the Ugric "7"], but the phonological details could perhaps be better explained by the assumption of a Proto-Tocharian origin"). Napol'skikh (1995: 124-125) has reconstructed the consonant stem \**Säpt* for the Ugric numeral "7", following Xelims'kij (1979: 121, 125). Also he prefers to see here a borrowing from ancestors of Tocharians. Proto-Tocharian \**šäpət* "7" (Winter 1992b: 109; see above) appears to be a more probable source of both the Ugric forms for "7" than Indo-Aryan \**sapta*. Concerning the other evidence of Tocharian-Ugric connections, cf. Ivanov about phonological parallelism (1986: 11-14) and Napol'skikh, summarizing the Tocharian - Fenno-Ugric parallels (1994: 37-39). He tries to identify the Tocharian influence with so-called Seima-Turbino archaeological phenomenon (17-16th cent. B.C.), deriving it from the Afanasievo culture (Napol'skikh 1994a), localized at Altai mountains beginning the 3rd mill. BC. (Mallory 1992: 62, 225).

c) Samoyed \**sejt<sup>3</sup>wê* (~ \**sejkwê* ?) "7" (Janhunen 1977: 139; *t<sup>3</sup>* = *c/k/s/t*)

Nganasan *šaibə*, *šaibúa*, Enets *se'o*, cf. Yurak (= early Enets) *tet-siù* "mensis" (4 x 7), Nenets (Tundra) *ši:β*, cf. ordinal *si'ivmdej*, (Forest) *še"eβ*;



Selkup *sel'či*; Kamassin *seigbu*, *sei'bu*, Koibal *sseigbe*, Mator *keipbe*, Taigi *kéibü*, Karagas *gydby*.

In spite of the incompatibility of inlaut consonantism, Honti (1993: 106), following the scholars as e.g. Castrén, Gombocz, Collinder, admits a relationship to Fenno-Permic *\*šejćcem*.

Janhunen (1983: 119) has modified the reconstruction in *\*sejptâ*, assuming a borrowing from proto-Tocharian. This solution is accepted by Napolskikh (1995: 119-121). He sees the most probable source in early Tocharian B, presenting a proper view on the phonetic development: B *šukt* < early B *\*šäwk(w)t3* > proto-Samoyed *\*sewktwê* > *\*sejktwê* > *\*sejkwê* / *\*sejtwê*. Again, the hypothetical contact of ancestors of Tocharians and Samoyeds can be localized in space and time. The dominant Tocharian ethnicity of creators of the Afanasievo culture occupying the territory between the upper Yenisei and the Altai mountains in the 3rd mill. B.C. (beginning even c 3500 B.C.) is usually accepted (Mallory 1995: 379-382). The most detailed overview of the facts localizing the proto-Samoyed homeland (3rd-1st mill. B.C.) was summarized by Xelimskij (1988: 13-14). He determines it by the territory between Ob and Yenisei, in the tetragon Naryn-Tomsk-Yeniseisk-Krasnoyarsk, inclusive North Altai and Sayany mountains. It means, that the bearers of the Afanasievo culture (= the ancestors of Tocharians ?) and the ancestors of Samoyeds were probably during the 3rd mill. B.C. neighbors. The Afanasievo culture was taken by turn the place of the Okunievo culture representing probably the Samoyed ethnos in the beginning of the 2nd mill. B.C. (Vadeckaja 1990: 73). Let us mention that the oldest Europoid mummies from Xinjiang in Northwest China (early Tocharians ?) are dated c 2000 B.C. (Mallory 1995: 381-382).

#### B. Kartvelian languages

Kartvelian *\*šwid-* "7" is reconstructed on the basis of (Old) Georgian *šwid-i*, Megrelian *škwit-i*, Laz *šk(w)it-i*, Swan *i-šgwid*, *i-šgüd*, ord. me: *-šgwde* (Klimov 1964: 216-217; Fähnrich and Sardshweladse 1995: 429). As early a writer as Bopp (see Klimov l.c.), reconstructing *\*šiwid-*, connected this numeral with Indo-European *\*septm̥*. Much more hopeful is the solution of Illič-Svityč (1964: 7; accepted by Gamkrelidze and Ivanov 1984: 875), who has found the most probable source in Semitic, cf. Akkadian *sibittu* "7" (see below). Klimov (1967: 308) accepts it. Later (1985: 206) he speculates about a modified bisyllabic archetype *\*šiwid-*.

Klimov (l.c.) has collected more words of Semitic origin in Kartvelian including numerals (besides "7" also "8", "9", "10"/"100"; Manaster Ramer 1995: 16-17 adds "5"). The Kartvelian-Semitic contact can be documented archaeologically as well. Safronov (1989: 242-258) has identified in the Maikop culture from northern Caucasus (26th-23rd cent. B.C.) genetic links to the Upper

Euphratian culture related to the Ebla civilization. Consequently he concludes that the bearers of the Maikop culture were Semites.

### C. Afroasiatic languages

a) Semitic \*šib-*u(m)* and \*šib-*át-u(m)* "7", formally m. and f. respectively, but in congruence they are used in the gender opposite to that of noun; this inversion of gender also operates when the numeral appears without an accompanying noun (Moscato 1964: 116). Dolgopolsky, an author of these reconstructions (p.c., Oct 1995), mentions that the feminine suffix is normally unaccented; he explains the function of the feminine-like marker \*-át- determining the Semitic numerals 3 - 10 and accompanying the masculine nouns as the original collective marker. The numeral continues in Old Akkadian *šabe*, later *sebe*, *seba* // *sebet(tum)*, *sibittu* etc., Ugaritic and Phoenician *šb* // *šb* *t*, Hebrew *šēba* // *šib* *a*:, Old Aramaic *šb* *t*, Jewish Aramaic *šəba* // *šab* *t* *o*:, Arabic *sab* *t* - // *sab* *t* *at* -, Sabeen *šb* *t* // *šb* *t*, Geez *sab* *t*, *səb* *t* // *sab* *t* *attu*:, Jibbali *so* *t* // *səb* *t* *et*, Harsusi *ho:ba* // *həbayt*, Mehri *ho:ba* // *yəbayt*, Soqotri *yhob* *t* // *hʷəb* *ah* etc. (Brugnatelli 1982; Dolgopolsky 1992: 34).

b) Egyptian *sḫw* // *sḫt* "7", m. // f. resp., vocalized \**safhaw* // \**safhat* after Middle Babylonian transcription *šap-ha* and Coptic (Ahminic) *sahf* // *sahfe*, (Sahidic) *sašf* // *sašfe* m. // f. (Vycichl 1983: 203). Egyptian *h* instead of expected *t* probably originated by alliteration to the following numeral *ḥmnw* // *ḥmnt* = \**ḥama:naw* // \**ḥama:nat* "8". One would expect the following spirantization \*-*bḫ*- > \*-*fḫ*-, but the cluster -*bḫ*- exists e.g. in *ḫbḫ* "to mix" or in *sbḫ.t* "a kind of amulet" (Vycichl 1983: 249, 185). Perhaps some combinatorical change has operated here; cf. the pair *ḫsf* vs. *ḫsb* "to succeed in protecting" (Edel 1955: 51). Vycichl (1983: 203) presents an alternative solution, assuming the following chain of substitutions: \*-*b* *t* - > \*-*b* *γ* - > \*-*f* *γ* - > \*-*fḫ*-. Finally Schenkel (1990: 56) sees in Egyptian *f* vs. Semitic \**b* regular reflexes of Afroasiatic \**p*; Egyptian *h* and Semitic \**t* have to reflect Afroasiatic \**γ* 1/\**γ* 2.

c) Berber \**sa:h<sub>2</sub>* (\**hissa:h<sub>2</sub>* ?) // \*-*at* "7", m. // f. (Prasse 1969: 19, 89; Id. 1974: 403, 405) > Ghadames *sa* // *sa:t*; Ghat *sa* // *sahət*, Ahaggar *əssa* // *əssahāt*, Ayr *əšša* // *əššayāt*, Awlimmiden *sah* // *sahat*; Zenaga *əššəh* // *əššəhət*; Mzab *sa* // *sa:t*, Semlal *sa* // *sāt* etc. and Guanche (Gran Canaria ?) *satti*, (Tenerife ?) *sa(t)*.

d) ? Chadic (Central): Gwendele, Hurzo *cíba* "7" (de Colombel; see Blažek 1990: 31). This Semitic-Egyptian-Berber(-Chadic ?) isogloss has probably no hopeful etymology within these language families, with the possible exception of Chadic, which may present a promising solution. The numeral "3" plays the key role here. There are two basic forms for the numeral "3" in Chadic:

(i) *\*kanu* and *\*kan(u)di* in West and Central branches; (ii) *\*suḫa* ~ *\*saḫu* in the Eastern branch: Mubi *súṣà*, Birgid *súúbù*, Jegu *sup* // *sub*, Migama *súbbà*, Dangla *súbbà*, Sokoro *sùbbá*, Tumak *sùb*, Ndam *sùp*, Sumrai *súbù*, Lele *sùbù*, Kabalai *sàp*, Kera *soope*, Kwang *suupáy* (Jungraithmayr and Ibrizsimow 1994: 327). And in some of these languages the numeral "7" is formed just through the mediation of the numeral "3": Sumrai (Nachtigal) *déna: súbù* "7" = *\*three [bent] fingers* (*dénum, dunum* "finger"), Ndam (Decorse) *wo subo* "7" = *woro* "4" + *supu* "3"; cf. also Tumak (Caprile) *ḡa:g-su:ùb* "7" : *su:ùb* "3", Gulei (Lukas) *dag suba* "7" : *cuba* "3", Miltu (Bruehl) *laksup* "7" : *sobo* "3". The glottalized *\*-ḫ-* (> Mubi -ḡ-) can regularly reflect the cluster *\*-bḫ-*. Thus, the Semitic-Egyptian-Berber (-Chadic) isogloss *\*sabḫ-u* // *\*sibḫ-u* "7" and the East Chadic numeral *\*suḫa* // *\*saḫu* "3" are fully compatible phonetically and semantically as well. The more primitive meaning of the East Chadic numeral "3" and the transparent structure of its derivative representing the numeral "7" allow us to conclude that the numeral "7" attested in Semitic, Egyptian, Berber and maybe, Chadic, is formed through the mediation of the numeral "3". It implies the following two patterns based on the numeral "3": (i) subtractive, i.e. "7" = "[10 -] 3" (cf. Sumrai above); (ii) additive, i.e. "7" = "[4 +] 3" or "3 [+ 4]" (cf. Ndam above and numerous other examples, e.g. in West Chadic: Gerka (Migeod) *praukum* "7" = *prau* "4" + *kun* "3" or Fyer (Jungraithmayr) *púrwon* "7" = *pít* "4" + *yoón* "3").

A similarity of Indo-European *\*séptm̥* "7" and esp. Semitic form *\*šibḫátum* "7" (with mimation expressing definiteness) is apparent. Already Møller (1909: 124) has connected these numerals (incl. the Egyptian counterpart), interpreting them as a common heritage. More recently, Bomhard and Kerns (1994: #188) and Bomhard (1996) reach the same conclusion. A more realistic solution seems to be a borrowing of the Semitic numeral into Indo-European: *\*šibḫátum* > *\*šíbḫatum* (after *\*šíbḫum*) > *\*šíbḫatum* > *\*séptm̥* (Illič-Svityč 1964: 7; Gamkrelidze and Ivanov 1984: 875; Dolgopolsky 1988: 16). Supported by other Indo-European words borrowed from Semitic it represents a strong argument for an early contact between these families. The most natural explanation seems to be a neighborhood of Semitic and Indo-European families, implicating the Near Eastern localization of the Indo-European homeland. Concerning the chronology,

this borrowing should precede the disintegration of Indo-European family, usually dated before 4000 B.C. (e.g. Mallory 1992: 127, 276 presents the estimation of the beginning of disintegration about 4500 B.C.).

#### D. Etruscan

Etruscan *semφ*-(-*ś*) "7" and *semφalχ*-(-*ls*) "70" (d'Aversa 1994: 47, 64) can be connected with the Indo-European or Semitic numeral "7". A borrowing is not excluded.

#### E. Basque

Basque *zazpi* /*saspil* "7" resembles very suggestively Coptic (Sahidic) *sašf*, *sašfe* (Bohairic) *šašf*, \**šašfi* m., f. "7" (von der Gabelentz 1894: 98-99; Löpelmann 1968: 1075). There are more lexical parallels between Basque and Coptic or late Egyptian collected esp. by von der Gabelentz (cf. Basque *sei* "6" vs. Coptic *sow* m., *soe* f. "6" ?). Any direct contact between Basque and Coptic // late Egyptian seems to be improbable. But the fact that in southern Spain some Egyptian hieroglyphic signs were discovered (Anderson 1988: 31) can support a certain kind of contact, perhaps mediated by Phoenicians.

### 4. Conclusion

The analyzed data can be summarized as follows:

- 1) East Chadic \**suḫa* ~ \**sabū* "3" can reflect Afro-Asiatic \**sab* ∩ -*u*.
- 2) Semitic-Egyptian-Berber(-Chadic) \**sab* ∩ -*u(m)* ~ \**sib* ∩ -*u(m)* "7" is probably formed through the mediation of the numeral "3", i.e. "7" = "[10 -] 3" ?
- 3) Semitic \**šib* ∩ *átum* "Siebenheit" was borrowed into Indo-European in the form \**séptm* "7".
- 4) Kartvelian \**šwid*- "7" was borrowed from a Semitic source close to Akkadian *sibittu* (Eblaic ?).
- 5) Fenno-Permic \**se(j)ććem* > \**šejććem* "7" was borrowed from a Baltic source close to Lithuanian *sėkmas* "7".
- 6) Ugric \**θäptε* and/or Mansi \**sä:tε* "7" were borrowed from Indo-Iranian \**sapta* or from proto-Tocharian \**śəpət*.
- 7) Samoyed \**sejptε* "7" was borrowed from proto-Tocharian \**śəpət*; the alternative reconstruction \**sejkwê* // \**sejtwê* indicates as the source some form preceding Tocharian B *šukt*.
- 8) Etruscan *semφ*- "7" could be borrowed from some Indo-European (Anatolian ?) or Semitic source.
- 9) Basque *zazpi* "7" was probably borrowed from a late Egyptian source close to Coptic (Sahidic) *sašfe*, (Bohairic) \**šašfi* f. "7".

### Postscriptum

The following analysis of the Indo-European numeral "7" has not previously been proposed. The cardinal *\*septm* is very difficult to analyze from our knowledge of Indo-European "Stammbildung." But this rather pessimistic conclusion is not quite valid for the ordinal *\*septmmo-*. Segmenting the numeral into *\*sep-* and *\*-tmmo-*, we can identify the latter member with the suffix of the superlative, reconstructed *\*-tmmo-*, (Brugmann) = *\*-tmo-* (Szemerényi) = *\*tmHo-* (Beekes). Accepting this point of view, it remains only for us to explain the function of the first component. There is essentially only one possible etymon in the Indo-European lexicon, *\*sep*, reconstructed on the basis of Old Indic *sap-* "pflegen, ehren, hochhalten, hegen", Avestan *hapti*: "beachtet, hält sich an, bewahrt", Greek *ἐπι* "besorge, betreibe, verrichte" etc. Pokorny (1959:909) assumes an original meaning *\*"sich in etwas abgeben, in Ehren halten."* This latter meaning may represent exactly a key to the semantic motivation of the numeral. The solution *\*septmmo-* = *\*"the most honorable"* corresponds fully to the prominent position of the numeral "7" among the Indo-Europeans. (This idea could be borrowed from the Semitic world.) The creation of the cardinal *\*septm* can be described as "ordinal" minus the "ordinal suffix, *\*(H)o-*," fully in agreement with the cardinal : ordinal opposition characterizing other numerals.

### References:

Abaev, Vasilij I.

- 1958 *Istoriko-ètimologičeskij slovar' osetinskogo jazyka, I.* Moskva, Leningrad: Izdatel'stvo Akademii nauk SSSR.

Abaev, Vasilij I.

- 1981 Doistorija indoirancev v svete arioural'skix jazykovyx kontaktov. In: *Ètničeskie problemy istorii Central'noj Azii v drevnosti (II tys. do n. è.)*, ed. M.S. Asimov, 84-89. Moskva: Nauka.

Anderson, James M.

- 1988 *Ancient Languages of the Hispanic Peninsula.* Lanham: University Press of America.

Beekes, Robert S.P.

- 1995 *Comparative Indo-European Linguistics: An Introduction.* Amsterdam, Philadelphia: John Benjamin.

Bereczki, Gábor

- 1992 *Grundzüge der Tscheremissischen Sprachgeschichte, II.* Szeged: Studia Uralo-Altaica 34.

Berger, Hermann

- 1986 Die Zahlwörter in den neuindoarischen Sprachen. *MSS* 47: 23-77.

Bernardo Stempel, Patricia de

- 1984 *Die Vertretung der indogermanischen liquiden und nasalen Sonanten im Keltischen.* Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft, Bd. 54.

Blažek, Václav

- 1990 A Comparative-Etymological Approach to Afrasian Numerals. In: *Proceedings of the Fifth International Hamito-Semitic Congress (Vienna, Sept 1987)*, ed. H.G. Mukarovsky, 29-44. Wien: Afro-Pub Beiträge zur Afrikanistik).
- 1990a New Fenno-Ugric - Indo-Iranian lexical parallels. In: *Uralo - Indogermanica (= UIG) II. Balto-slavjanskije jazyki i problema uralo - indo-evropejskix svjazej (Materials of the conference held in Moscow, June 18-22, 1990)*, eds. V.V. Ivanov, T.M. Sudnik, E.A. Xelimskij, 40- 43. Moskva: Institut slavjanovedenija i balkanistiki.

Bomhard, Allan R.

- 1996 *Indo-European and the Nostratic Hypothesis.* Charleston: Signum (Studia Nostratica, 1).

Bomhard, Allan R. and John C. Kerns.

- 1994 *The Nostratic Macrofamily: A Study in Distant Linguistic Relationship.* Berlin: Mouton de Gruyter.

Brugmann, Karl

- 1892 *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen, II.2.* Strassburg: Karl J. Trübner.

Brugnatelli, Vermondo

- 1982 *Questioni di morfologia e sintassi dei numerali cardinali semitici.* Firenze: La nuova Italia editrice.

Coleman, Robert

- 1992 Italic [numerals]. In: *IEN*, 389-445.

Comrie, Bernard

- 1992 Balto-Slavonic [numerals]. In: *IEN*, 717-833.

D'Aversa, Arnaldo

1994 *Dizionario della lingua etrusca*. Brescia: Paideia Editrice.

Debrunner, Albert and Wackernagel, Jacob

1930 *Altindische Grammatik, III: Nominalflexion - Zahlwort - Pronomen*. Göttingen: Vandhoeck und Ruprecht.

Dolgopolsky, Aron

1988 The Indo-European and lexical contacts of Proto-Indo-European with other languages. *Mediterranean Language Review* 3: 7-31.

1992 *From Proto-Semitic to Hebrew*. Phonology. Milano: Studi camito-semitici (forthcoming).

Edel, Elmar

1955 *Altägyptische Grammatik*. Roma: Pontificium Institutum Biblicum.

Eichner, Heiner

1992 Anatolian [numerals]. In: *IEN*, 29-96.

Emmerick, Ronald

1992a Old Indic [numerals]. In: *IEN*, 163-197.

1992b Iranian [numerals]. In: *IEN*, 289-345.

Fährnrich, Heinz and Surab Sarashweladse

1995 *Etymologisches Wörterbuch der Kartwel-Sprachen*. Leiden: E.J. Brill (Handbuch der Orientalistik, Abt. 1, Bd. 24).

Fraenkel, Ernst

1962-65 *Litauisches etymologisches Wörterbuch, I-II*. Göttingen: Vandhoeck and Ruprecht / Heidelberg: Carl Winter.

Gamkrelidze, Tamas V. and Ivanov, Vjačeslav V.

1984 *Indoevropskij jazyk i indoevropcjy*. Tbilisi: Izdatel'stvo tbilisskogo univerziteta.

Georgiev, Vladimir I.

1977 *Trakite i texnijat ezik*. Sofija: Izdatelstvo na Balgarskata akademija na naukite.

Gordeev, Fedor I.

1985 *Istoričeskoe razvitie leksiki marijskogo jazyka*. Joškar-Ola: Marijskoe knižnoe izdatel'stvo.

Greene, David

1992 Celtic [numerals]. In: *IEN*, 497-554.

Gvozdanovič, Jadranka, ed.

1992 *Indo-European Numerals*. Berlin / New York: Mouton de Gruyter.

Hamp, Eric

1952 The anomaly of Gmc. '7'. *Word* 8: 136-39.

1992 Albanian [numerals]. In: *IEN*, 835-921.

Honti, László

1982 *Geschichte des obugrischen Vokalismus der ersten Silbe*. Budapest: Akadémiai Kiadó.

1993 *Die Grundzahlwörter der uralischen Sprachen*. Budapest: Akadémiai Kiadó.

*IEN* see Gvozdanovič 1992.

Illič-Svityč, Vladislav M.

1964 Drevnejšie indoevropsko-semitskie jazykovye kontakty. In: *Problemy indoevropskogo jazykoznanija*, ed. V.N. Toporov, 3-12. Moskva: Nauka.

Ivanov, Vjačeslav V.

1986 Parallelizm fonologičeskix sistem ranneugorskogo i obščetoxarskogo prajazykov i ego vozmožnoe diaxroničeskoe objasnenie. In: *Fonetika jazykov Sibiri i sopredel'nyx regionov*, ed. V.M. Nadeljaev, 11-14. Novosibirsk: Nauka.

Janhunen, Juha

1977 *Samojedischer Wörschatz. Gemeinsamojedische Etymologien*. Helsinki: Castrenianumin toimitteita 17.

1983 On early Indo-European - Samoyed contacts. In: *Symposium Saeculare Societatis Fenno-Ugricae*, 115-127. Helsinki: MSFOu 185.

Joki, Aulis

1973 *Uralier und Indogermanen*. Helsinki: MSFOu:151.

Jungraithmayr, Herrmann and Ibriszimow, Dymitr

1994 *Chadic Lexical Roots, II. Documentation*. Berlin: Dietrich Reimer.

Keresztes, László

1986 *Geschichte des Mordwinischen Konsonantismus, II*. Szeged: Studia Uralo-Altaica 26.



Klimov, G.A.

- 1964 *Ėtimologičeskij slovar' kartvel'skix jazykov*. Moskva: Nauka.
- 1967 Zaimstvovannye čislitel'nye v obščekartvel'skom. *Ėtimologija* 965: 307-310.
- 1985 Zu den indogermanischen-semitisch-kartwelischen Kontakten im Vorderen Asien. In: *Sprachwissenschaftliche Forschungen*. Fs. für J. Knobloch, eds. M. M. Ölberg and G. Schmidt, 205-210.

Korenych, Éva

- 1972 *Iranische Lehnwörter in den obugrischen Sprachen*. Budapest: Akadémiai Kiadó.

Kortlandt, Frederik

- 1983 Greek numerals and PIE glottalic consonants. *MSS* 42: 97-104.
- 1994 Proto-Armenian numerals. In: *In honorem Holger Pedersen. Kolloquium der Indogermanischen Gesellschaft (Kobenhagen, März 1993)*, ed. Jens E. Rasmussen, 253-257. Wiesbaden: Ludwig Reichert Verlag.

Kronasser, Heinz

- 1966 *Etymologie der hethitischen Sprache, Band 1*. Wiesbaden: Otto Harrassowitz.

Kuz'mina, Elena E.

- 1994 *Otkuda prišli indoarii*. Moskva: Vostočnaja literatura.

Lamprecht, Arno T.

- 1987 *Praslovanština*. Brno: Univerzita J.E. Purkyně.

Lehmann, Winfred P.

- 1986 *A Gothic Etymological Dictionary*. Leiden: E.J. Brill.

Lehtiranta, Juhani

- 1989 *Yhteissaamelainen sanasto*. Helsinki: MSFOu:200.

Lewis, Henry and Pedersen, Holger

- 1954 [37] *Kratkaja sravnitel'naja grammatika kel'tskaja jazykov*. Moskva: Izdatel'stvo inostrannoju literatury [Russian translation of *A Concise Comparative Celtic Grammar*. Göttingen].

Löpelmann, Martin

- 1968 *Etymologisches Wörterbuch der baskischen Sprache*. Berlin: Walter de Gruyter.

Lytkin, Vasilij I. and Guljaev, Evgenij S.

1970 *Kratkij ètimologičeskij slovar' komi jazyka*. Moskva: Nauka.

Mallory, James P.

1992 *In Search of the Indo-Europeans: Language, Archaeology and Myth*. London: Thames and Hudson [reprint of the first edition of 1989].

1995 Speculations on the Xinjang Mummies. *JIES* 23: 371-384.

Manaster Ramer, Alexis

1995 Some Borrowed Numerals in Proto-Kartvelian. *Dhumbadji* 2/3:16-18.

Mann, Stuart

1977 *An Albanian historical grammar*. Hamburg: Buske.

1984-87 *An Indo-European Comparative Dictionary*. Hamburg: Buske.

Mayrhofer, Manfred

1976 *Kurzgefasstes etymologisches Wörterbuch des Altindischen, Band III*. Heidelberg: Carl Winter.

1996 *Etymologisches Wörterbuch des Altindoarischen, II. Lieferung* 19. Heidelberg: Carl Winter.

Melchert, H. Craig

1993 *Cuneiform Luvian Lexicon*. Chapel Hill: Lexica Anatolica 2.

Morgenstierne, Georg

1927 *An Etymological Vocabulary of Pashto*. Oslo: Jacob Dybwad.

Moscatti, Sabatino et al.

1964 *An Introduction to the Comparative Grammar of the Semitic Languages. Phonology and Morphology*. Wiesbaden: Otto Harrassowitz.

Møller, Hermann

1909 *Indoeuropæisk-semitisk sammenlignende glossarium*. Kjøbenhavn: J.H. Schultz.

Napolskikh [Napol'skix], Vladimir V.

1994 O vremeni i istoričeskix uslovijax uralo-toxarskix kontaktov. *JSFOu* 85: 37-39.

1994a Uralic and Tokharian: Linguistic evidence and archaeological data. Paper presented at the 3rd World Archaeological Congress, New Delhi, Dec 1994.

1995 URALIC "seven". *JSFOu* 86: 119-128.

Parpola, Asko

1994 *Deciphering the Indus Script*. Cambridge: University Press.

Prasse, Karl-G.

1969 Á propos de l'origine de l'h touareg. København: *Det Kgl. Danske Videnskabernes Selskab, Hist.-filosof. Medd.* 43.3.

1974 *Manuel de grammaire touaregue (tāhāggart), IV-V*. Nom. Copenhagen: Akademisk Forlag.

Rédei, Károly, ed.

1988 *Uralisches etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.

Ross, Alan S.

1941 Some remarks on the numerals of Finno-Ugrian. *Transactions of the Philological Society* 1941: 1-15.

Ross, Alan S. and Berns, Jan

1992 Germanic [numerals]. In: *IEN*. Pp. 555-715.

Safronov, Vladimir A.

1989 *Indoevropskie prarodiny*. Gorki: Volgo-Vjatka Publishing House.

Sammallahti, Pekka

1988 Historical Phonology of the Uralic Languages. In: *The Uralic Languages*, ed. D. Sinor, 478-554. Leiden: E.J. Brill.

Schenkel, Wolfgang

1990 *Einführung in die altägyptische Sprachwissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Schmid, Wolfgang P.

1989 *Wort und Zahl. Sprachwissenschaftliche Betrachtungen der Kardinalzahlen*. Wiesbaden: Franz Steiner.

Schmidt, Gernot

1992 Indogermanische Ordinalzahlen. *IF* 97:197-235.

Schwyzer, Eduard

1939 *Griechische Grammatik, Band 1*. München: C.H. Beck.

Sedov, V.V.

1994 Vostočnoslavjanskaja ètnojazykovaja obščnost'. *Voprosy jazykoznanija* 1994/4: 3-16.

Serebrennikov, B.A.

1963 *Istoričeskaja morfologija permskix jazykov*. Moskva: Izdatel'stvo AN SSSR.

Shevoroshkin, Vitaly

1979 On the Hittite-Luwian Numerals. *JIES* 7: 177-198.

Shields, Kenneth

1992 On the origin of the Germanic decads, 70- 100. *NOWELE* 19: 89-100.

Smoczyński, Wojciech

1989 *Studia bałto-słowiańskie, I*. Kraków: Ossolineum (Prace komisji językoznawstwa, Nr. 57).

Stang, Christian S.

1966 *Vergleichende Grammatik der Baltischen Sprachen*. Oslo, Bergen, Tromsø: Universitetsforlaget.

Szemerényi, Oswald

1960 *Studies in the Indo-European system of numerals*. Heidelberg: Carl Winter.

Thurneysen, Rudolf

1946 *A Grammar of Old Irish*. Dublin: Institute for Advanced Studies.

Tkačenko, Orest B.

1989 *Očerki teorii jazykogo substrata*. Kijev: Naukova dumka.

UIG see Bláček 1990a.

Vadeckaja, È. B.

1990 Južno-samodijskie komponenty kul'tury drevnego naselenija Prisajan'ja. In: *UIG* II, 71-80.

Van Windekens, Albert J.

1976 *Le Tokharien confronté avec les autres langues indo-européennes, I. La phonétique et la vocabulaire*. Louvain: Centre International de Dialectologie Générale.

von der Gabelentz, Georg

- 1894 *Verwandschaft des Baskischen mit den Berbersprachen Nord-Africas*.  
Braunschweig: Richard Sattler.

Voyles, Joseph

- 1987 The cardinal numerals in Pre- and Proto-Germanic. *Journal of English and Germanic Philology* 86: 487-495.

Vycichl, Werner

- 1983 *Dictionnaire étymologique de la langue copte*. Leuven: Peeters.

Waanders, Frederik M.J.

- 1992 Greek [numerals]. In: *IEN*, 369-388.

Winter, Werner

- 1992a Some thoughts about Indo-European numerals. In: *IEN*, 11-28.  
1992b Tocharian [numerals]. In: *IEN*, 97-161.  
1992c Armenian [numerals]. In: *IEN*, 347-359.

Xelimskij, Evgenij A.

- 1979 Čeredovanie dolgot, konsonantnyj auslaut i udarenie v istorii  
vengerskix imennyx osnov. In: *Balcanica*, 118-133. Moskva: Nauka.  
1988 *Istoričeskaja i opisatel'naja dialektologija samodijskix jazykov*. Tartu:  
Dissertacija.

Zaliznjak, A.A.

- 1988 Drevnenovgorodskij dialekt i problemy dialektного členenija  
pozdnego praslavjanskogo jazyka. In: *Slavjanskoe jazykoznanie. X  
Meždunarodnyj s'ezd slavistov*, ed. N.I. Tolstoj, 164-177. Moskva:  
Nauka.

Zinkevičius, Z.

- 1984 Pol'sko-jatvjazskij slovarik? *Balto- slavjanskie issledovanija* 1983: 3-  
30.

# The Phonotactics of Sumerian

Claude Pierre Boisson  
*Université Lumière, Lyon*

## Introduction

In this paper I shall examine two aspects of Sumerian phonotactics, the problem of syllable structure (specifically tautosyllabic consonant clusters) and the problem of vowel harmony. I shall do this from the angle of typology and the universals, that is, I shall apply the findings of general linguistics to Sumerian phonology, a method which has hardly ever been used so far<sup>1</sup>. In so doing, I shall neither follow those who contend that the notion of phonological system cannot be applied to dead languages, nor those who are willing to apply it, but without any consideration of the phonetic realizations — these views are expressed respectively by Laroche (1960: 259) and Sollberger (1950: 54). My own position on the use of phonology is basically that of Lupaş on Greek and the authors quoted by her (Lupaş 1972: 54-56). It is exemplified in a number of papers of mine on proto-languages (Indo-European) or on dead languages such as Sumerian, Etruscan and Carian (Boisson 1989a, 1989c, 1991a, 1991b, 1994).

At the outset, a few specifications on the notational conventions are in order. I shall use the (personal) coinage "translitereme", which is convenient to avoid any ambiguity. Strictly speaking, "transliteremes" are the traditional (and phonologically unrealistic) Assyriological notation for individual graphemic units, which are to be distinguished both from the cuneiform units, the cuneograms (which are

<sup>1</sup> I wish to express my sincere thanks to Igor Diakonoff, Dietz Otto Edzard, Fernande Krier, Gilbert Puech, and Robert Vago for various pieces of information and various remarks. None of them should be held responsible for my failings. The present paper is an excerpt from a long paper on Sumerian phonology in which the vowels are examined (Boisson 1991b).

the real graphemes for Sumerian), and from the phonological entities, the phonemes (to say nothing of phones, the phonetic units). Strictly defined, the transliteremes are indivisible units in one-to-one correspondence with the cuneograms. In conformity with standard linguistic notation, I shall use angled brackets (< >) for transliteremes, but they have a different function in Assyriological literature. They will also be used more loosely, but will be distinguished from the symbols //, which refer to phonemes, and [ ], which refer to speech-sounds ("phones"). So, in order to avoid ambiguity, heterogeneity of notations, as well as terminological and conceptual confusions, a situation which is sometimes encountered in Assyriological writings, we should carefully distinguish the following notations for e.g. "his king": (1) graphemic transliteration, < lugal-a-ni >; (2) morphemic transcription { lu+gal-ani }; (3) phonemic (phonological) representation /lukalani/; (4) phonetic representation [lukalani] or [lugalani] (on all this see Boisson 1989a). We should painstakingly avoid mixing these representations. Needless to say, moving from (1) to (4) means going from the most assured to the most hypothetical symbolization.

### **Tautosyllabic consonant clusters**

To my knowledge, the issue of the syllabic structure in Sumerian has never been examined systematically, so that the following discussion on tautosyllabic consonant clusters may prove useful as a starting-point.

Syllabography does not favour the notation of tautosyllabic consonant clusters, whether syllable-final or syllable-initial. This is obvious from the examination of strict syllabaries, i.e. excluding mixed systems like the Old Persian cuneiform or the Iberian scripts. To my knowledge, only the Cherokee syllabary comes close to providing a clear counter-example, since out of a total of 85 signs it includes no fewer than 20 signs in systematic use for /CCV-/ syllables. Some syllabaries include only a handful of signs for syllables with consonant clusters: Linear B has a small number of optional

signs for the configuration consonant + semi-vowel + vowel<sup>2</sup>; we also find 2 signs for /ksa/ and /llam/ in the Siddham system used in China and Japan to transcribe Sanskrit (cf. the *Pumsu* script in Korea). These are to be treated as merely marginal facts, just as the existence of syllabic signs in the predominantly alphabetic systems of the Coptic, Meroitic, Glagolitic or Cyrillic scripts. But predominantly, syllabaries are adapted to languages of the /CV/ (phonological) syllable type, such as Japanese (even though Japanese may have phonetic clusters, which is another matter). When they are used for languages with phonemic consonant clusters, such as the Indo-European languages, they create a graphic image which cannot do justice to the real pronunciation, as is only too obvious in the treatment of Greek in Linear B and the Cypriote syllabary, or in the distortions imposed by the cuneiform syllabary on Hittite or the "Hittite hieroglyphic" syllabary on Luwian. These Anatolian languages certainly had tautosyllabic consonant clusters, like all Indo-European languages, as this is proved by their existence in more

2 Out of 103 clearly identified signs, Linear B has 16 supplementary signs, among which are found the following: < twe >, < two >, < dwe >, < dwo >, < tja >, < nwa >, < rja >, < rjo >, < pte >, and perhaps < swa > and < swi >. There are two remarkable properties of these signs. The first is that, with the apparent exception of < pte >, all of them evince the same pattern and are used to note a /Cw/ or /Cj/ cluster, i. e. a cluster which has a special status (and which may correspond to a phonemic characteristic of the Minoan, Eteocretan language for which the ancestral form of the syllabary was devised, as hypothesized by L. R. Palmer). Even < pte > can be brought into line, as Palmer has suggested that it comes from a syllabogram originally devised for \* < pje >; he also thinks that the much-discussed < zV > signs correspond to palatalized /kjV/ syllables. The second fact is that these signs are of optional use, and that we have graphic alternants such as < pte-re-wa > ~ < pe-te-re-wa >. In other terms the Linear B script does not offer a fully integrated treatment of consonant clusters, as against the Cherokee syllabary. On Linear B see: Chadwick 1987; Doria 1968; Lejeune 1966; Palmer 1963; Palmer 1980; Stephens & Justeson 1978; Ventris & Chadwick 1973.



recently attested languages of the same family that were written alphabetically (Lycian, Lydian).

True, the cuneiform syllabaries have the almost unique property of being capable of expressing closed syllables, since they have signs for various <(C)V(C)> combinations — the Chinese system being apparently the only other similar system<sup>3</sup>. This enables them to express heterosyllabic consonant clusters without any problem, as in Akkadian /wardum/ "servant" noted <wa-ar-du-um>. But tautosyllabic clusters cannot be expressed directly in this system. This is fine for Akkadian, which, like most other Semitic, or even most Afroasiatic, languages, does not know them. But this poses a problem for other languages. We can probably trust a notation such as <anše> "donkey", and surmise that Sumerian had heterosyllabic clusters. But if Sumerian had tautosyllabic clusters, there is no simple way to tell from the script. Hittite /tri/ "three" written in the cuneiform syllabary comes out as <te-ri> or <ta-ri>. In Mycenaean Greek an initial consonant cluster is written as a series of open graphic syllables ending with the same vowels, while /s+C+V/ is written <CV->, even though five <sV> syllabograms were available to do the job approximately. Incidentally, this ellipsis of /s/ in the script might be accounted for by the fact that in a number of languages, /sC/ clusters seem to have special properties, /s/ and the following /C/ being felt as more closely bound than the members of other clusters: so much so that a number of phoneticians after Firth have treated them as a functional unit (see Davidsen-Nielsen 1974; Doria 1968; Ewen 1982; Fudge 1969); and the reverse pattern may be treated similarly, as is shown by the fact that the Milesian alphabet of Greek has

<sup>3</sup> The possibility of noting /VC/ is particularly original. The only similar case I can think of is the "runic" script of the Orkhon, where this is systematic — this writing was used for Old Turkish, a (C)(C)V(C)(C) language. It is also significant that the names of the letters in the Nikolsburg alphabet for Hungarian are of the type /VC/ — ours being mostly /CV/, except for the names of <l>, <m>, <n>, <r>, taken over from the Etruscan names for these consonants, which could be syllabic.

special letters for /ks/, /ps/<sup>4</sup>. Thus the words /kla:re:wes/, /sperma/ come out as <ka-ra-re-we>, <pe-ma> — and not <se-pe-ma> or <se-pe-re-ma>, theoretical possibilities allowed by the existence of the required syllabograms, which would have been better solutions in a way, similar to those encountered in the Cypriote script.

The rationale of such systems poses special problems for the Sumerologists. Those who think that the fit between writing and pronunciation is not too inadequate may take at their face value notations such as <su-kud> "high", <bu-lu-ug> "boundary", <ku-ru-un> "wine". But is such optimism totally justified? If the cuneiform syllabary worked more or less like the Linear B script, some of the writings might well hide the existence of tautosyllabic consonant clusters.

Just for the sake of the argument, let us imagine that Sumerian had clusters such as /sp-, st-, sk-, pl-, pr-/ or /-lp-, -lt-, -rt-, -rk/, etc. How would a scribe go about writing them? If he wants to write, say, something like /spa/, there are two solutions available to him: either omit the initial /s-/ altogether and write just <ba>, or break up the cluster and write <sa-ba>. If he wants to write /tra/, a solution can be <da-ra>. The situation is basically the graphic counterpart of what happens in phonology when Indo-European words are borrowed in Japanese, or again in Hebrew, Aramaic or Arabic (where the original pattern /C<sub>1</sub>C<sub>2</sub>V/ is naturalized as /C<sub>1</sub>V.C<sub>2</sub>V/ with an epenthesis and without any cluster at all,

<sup>4</sup> "On peut se demander si [psi et ksi] ne notaient pas, à l'origine, des phonèmes uniques dans le dialecte ionien" (Lupaş 1972: 108). Of course these clusters have the property of being found both word-initially and word-finally. Similarly, note that in those basically alphabetic scripts that have a few signs for consonant clusters, these clusters usually evince some special property that may account for the feeling of a particularly close connection of the constituting consonants, as is shown by the examination of the Orkhon script of the Old Turkish inscriptions in Siberia and Mongolia, the Nikolsburg alphabet of Hungarian, the Elbasaran script for Albanian, the Old Rumanian script, or the Bugis script (Sulawesi); and, significantly, in the latter case, the four letters for clusters are in optional use.

or /? V C<sub>1</sub>.C<sub>2</sub>V/ with a prosthesis and permissible heterosyllabic cluster). The consequence is that one moves easily from the pronunciation to the syllabic script, but the reverse process cannot avoid irrecoverable loss of information for Sumerologists.

What this implies is that some words written <CV> may actually correspond to <CCV>, and also that some words written as graphic disyllabic items can in fact bear multiple phonemic interpretations. Let us take the latter case, that of a <CVCVC> word, particularly (but not exclusively) those of the type <CVr/IVC>, where C's are obstruents, particularly stops. Given the constraints of syllabographic notation, there are no fewer than 8 a-priori phonological interpretations for it: (1) <CVCVC> does correspond to a phonological /CVCVC/ structure; (2a) <CVCVC> corresponds to a monosyllable with prevocalic /CC/ in the onset of the syllable, /CCVC/; (2b) <CVCVC> corresponds to a monosyllable with postvocalic /CC/ in the coda of the syllable, /CVCC/; (2c) <CVCVC> corresponds to a word with a consonantal peak, /CCC/; (3) if the Linear B-type graphic "loss" of /s/ obtains, <CVCVC> may correspond to a phonological structure with initial /sC-/ , with the sub-patterns (3a) /sCVCVC/; (3b) /sCVCC/; (3c) /sCCVC/; (3d) /sCCC/. Of course we must regard the writing system as coherent, so that one has to choose between /sCV/- --> <sVCV> (the Cypriote solution) and /sCV/- --> <CV-> (the Linear B solution), the two treatments of the /sC/ cluster being incompatible.

For instance take <gu-ru-un> = <gurun> "fruit". At least theoretically it may hide either: (1) /kurun/; or (2) /krun/, realized [krun], or even [kurun] with a purely phonetic svarabhakti; or (3) /kurn/, realized [kurn], or even [kurun] with a purely phonetic svarabhakti; or (4) /krn/, realized [kr̥n] with a syllabic [r̥], perhaps with a very elusive transition vowel; or (5) /skurun/; or (6) /skurn/; or (7) /skrun/; or even (8) /skrn/ realized [skr̥n] (cf. the Czech word for "death" below).

The solutions with the syllabic consonant may strike some readers as highly improbable and perhaps even fanciful, but Sumerologists should not unduly restrict the phonological possibilities for Sumerian, as they must realize that syllabic consonants, as in Czech <krk> "neck", <smrt>

"death" are extremely widespread — so much so that they are encountered in one third of the languages in the large sample collected by Hagège (1982: 21). They include familiar languages such as Amharic, Moroccan Arabic, Berber, Czech, English, Japanese, Khmer, various Sino-Tibetan languages including Mandarin Chinese, and Sanskrit; syllabic consonants have also been posited with a high degree of likelihood in Proto-Indo-European, Coptic, Etruscan. As jocularly pointed out by Walker (1987: 12): "No Sumerian cartoonist could write 'Psst!'"'. So when we read <su-kud>, <bu-lu-ug>, how are we to know for sure whether they correspond to /suk<sup>h</sup>ut/, /puluk/, or to /sk<sup>h</sup>ut/, /pluk/, or /suk<sup>h</sup>t/, /pulk/, or even /spulk/, etc. ?

Indeed there may be some vague clues as to the existence of consonant clusters. For instance <kalag> "strong" can be written <kal-ga>, which might perhaps suggest /k<sup>h</sup>alk/. Indeed a number of authors, such as Poebel and Jestin believed that Sumerian had consonant clusters (Jestin 1951: 35-36; Jestin 1954: 18-20), such as <gurd> ~ <gurud> "to throw", <kalg> ~ <kalag> "powerful", <bulg> ~ <bulug> "to walk", <gurs> ~ <gurs> "youth", etc. The sign KEŠ "to bind" has been interpreted as what Jestin writes "kešd" and Thomsen (1984) writes /kešdr/ (but Thomsen does not use // strictly for a phonemic notation), i.e. perhaps /k<sup>h</sup>eʃt<sup>r</sup>/, /t<sup>r</sup>/ being of course an affricate, a unitary phoneme, and not /t+r/.

Admittedly, there is an alternative interpretation for such facts. Falkenstein views the alternation <kalaga> ~ <kalga> as the expression of the syncope of an unstressed vowel, stress falling on the first syllable. He offers the same interpretation for the fact that <sikil> "pure", seems to be <skil> in <ki-sikil> "pure place = maiden", written <ki-iš-ki-il>. To this must be added that we read the late Greek transliteration <ΚΙCΚΙ[λ> on a tablet (B.2 = B.M. 34816, Sollberger 1962). The compound would be fore-stressed, so that instead of <ki-sikil> we would have <ki-skil>. Note however that such graphic alternations can be interpreted in two ways: either as the faithful rendering of a vowel syncope under weak stress, or as a purely artefactual consequence of syllabographic constraints. Suppose that

"pure" always was /skil/ in the first place. The optimal rendition by a scribe would be <si-ki-il> (the approximation <iš-ki-il> with graphic prosthesis being excluded by the rules of the system, at least word-initially), and this would create for us the illusory interpretation /sikal/. Now when the word is tagged at the end of <ki> as the second element of a compound, given the availability of <VC> syllabograms, the writing <ki-iš-ki-il> surfaces automatically as the best solution. The <sikal> ~ <skil> alternation would then be strictly graphic; indeed one must realize that it would not occur if the syllabary was of a Cypriote type, with only <CV> syllabograms, as the word "pure" would then have to be written something like <si-ki-le> in all cases.

An objection to the existence of tautosyllabic consonant clusters in Sumerian might take the following form: the cuneiform syllabary was first devised by Sumerian speakers for Sumerian, and must have been adequate to that language. This would make the situation significantly different from that of Linear B or the Cypriote syllabary, since these were evolved from scripts that were not created for Greek in the first place, so that the adaptation was far from optimal. But note that the Luwian syllabary had no <CCV> sign, even though Luwian, as an Anatolian language, most probably had /CCV/ configurations. Nevertheless the Luwian logo-syllabic writing seems to have been designed specifically for that language, probably with the impact of "stimulus diffusion", as pointed out by Meriggi. So we must not take it as a proven fact that the Sumerian syllabic notation was perfectly adequate to the Sumerian language, even if it was devised by the Sumerians themselves. In any case, if Sumerian had clusters, it would be expecting too much if we insisted that Sumerians must have given a faithful material image of the clusters at a stage of linguistic and graphic thinking where consonants probably could not be considered in themselves, but only in their association with a vocalic accompaniment — cf. the name of our own letters in spelling, where <p> is called /pi:/, /pe/, /pə/, and not just /p/: see Durand (1977: 45). Even in the sophisticated linguistic thinking of the Indians, the basic element was taken to be the syllable (akṣara), a term which was also used to refer to the individual

characters of the scripts, where, as is well known, the basic unit noted /Ca/, and not just /C/.

In any case, we must recall that if the Sumerian writing system had been adequate to the language from the point of view of syllabic structure, we would have trouble explaining why it was inadequate from the point of view of the vowel system, since it failed to provide for the very likely /o/ phoneme.

Perhaps in a few favourable cases the treatment of words borrowed from Sumerian into Akkadian might provide clues as to the existence of clusters, but the phonotactic patterns of Akkadian are such that the outcome is likely to hide original clusters too.

So if we do not jump to premature conclusions because of the script, or because of the habits ingrained in the work of Semitists, we may want to leave open a number of options for the syllabic structure of Sumerian. It may be a (C)V(C) language as traditionally supposed<sup>5</sup>, just like, say, Ainu, Araucanian, Buryat, Crow, Jivaro, Mandarin, Nahuatl, Quechua, or Dravidian languages. Or it may be a (C)V(C)(C) language like many Turkic, Finno-Ugric, Northeast Caucasian and Australian languages. Or a (C)(C)V(C) language like Chinantec, Gilyak, Inupik, Island Carib, Kiowa Apache and Malinka. Or it may even be a (C)(C)V(C)(C) language like Aleut, Baluchi, Chukchi, Ket and Wakhi.

We should also have an integrated view of the syllabic system of Sumerian, so as to avoid incompatibilities. Phonological forms with an uncommon word-initial consonant cluster are postulated by Civil (1982: 10), who suggests \*/lgud<sup>1</sup>/ for what is usually read as <lugud<sub>2</sub>> "short, thin". Such an initial cluster may seem odd, but is not unthinkable from a cross-linguistic point of view, since we find a similar one in various languages, for instance in Bahnar, a Mon-Khmer language of Vietnam, in the word *lpiet* ("language") (Hagège & Haudricourt 1978: 88, note). But we

<sup>5</sup> Here we are considering the syllable structure of roots, since some monomorphemic grammatical morphemes are realized as just /C/: the pronominal verbal prefix <-n-> (3 sg. animate) and <-b-> (inanimate) and the "conjugation prefix" <-m-> ("ventive"?).

should not take Civil's suggestion by itself. If there is indeed an initial phonemic cluster consisting of a liquid followed by a stop, so that the word would be something like /lkut/ (rather than /lgud<sup>r</sup>/), then we cannot accept this as an isolated hypothesis, but as part of a pair of interconnected phonotactic constraints (the necessity of taking a global, systemic view of Sumerian phonology, is a leitmotiv of my paper "Contraintes typologiques..."). In this connection we must bear in mind Greenberg's universal: "In initial systems [= consonant clusters, C.B.] the existence of at least one sequence containing a liquid, whether voiced or unvoiced, immediately followed by an obstruent [= stop, affricate, or fricative, C. B.] implies the existence of at least one sequence containing an obstruent immediately followed by a liquid" (Greenberg 1978: 258). This constraint means that if /lk-/ existed in Sumerian, then it must have known at least one of initial clusters such as /pl-/ , /pr-/ , /tl-/ , /tr-/ , /kl-/ , /kr-/ , /sl-/ , /sr-/ and suchlike. Languages of this phonotactic type in Greenberg's sample are: Aguacatec Mayan, Balti (Tibetan), Bilaan (Philippines), Chatino (Otomanguean), Coos (Penutian), Czech, Georgian, Khasi (Mon-Khmer), Pashto, Polish, Russian, Mitla Zapotec, i.e. 12 languages in a sample of 104, out of which 90 have initial consonant clusters.

### **Vowel harmony**

Sumerian had a clear tendency to vowel harmony (Falkenstein 1959a, 1959b; Krecher 1969; Diakonoff 1983: 87). Part of it is the type of harmony shown in prefixes (across morpheme boundaries), as studied by Poebel (1931) and Kramer (1936), and part of it is the type of harmony that appears within roots. The latter is a harmony of a very special type indeed (a fact that Assyriologists should bear in mind), which could be called "total vowel harmony" ("totale Vokalangleichung"). This is because Sumerian disyllabic terms have a strong tendency to show exactly the same vowel in the two syllables, instances being: <uru> "city", <gibil>

"new", <eme> "tongue", <amar> "calf", or <eze> in the Emesal dialect<sup>6</sup>.

Now Diakonoff takes this fact as the effect of an absolute rule. So that any apparent violation will have to be accounted for in two ways: (a) either the term is a compound, such as <lugal> "king", formed on <lu!> "human being" + <gal> "big"; (b) or it is a loanword, as for instance <apin> "plough", which Landsberger had already considered as borrowed by the Sumerians from an otherwise undetermined "Pre-Euphratic" people which was already in Mesopotamia when the Sumerians settled there in the 4th millennium, a statement which has continually been repeated since<sup>7</sup>.

But in such a reasoning I see risks of *circularity*<sup>8</sup>. If we decide that total vowel harmony is not simply a strong tendency, but an absolute rule that applies without any exception to any non-compound word of the indigenous Sumerian stock, then obviously any term violating this rule will be branded as a loan, which threatens to simplify excessively the formidable problem of pre-Sumerian substrates. There is no question that there exists a strong tendency to homogenize the vowels of disyllables, *but* some words may have had an older *non* homogenous form. For instance Edzard mentioned to me for <g\$urus&> ("a youth") the discovery at Ebla of a sign name <NU-rí-su>, where <NU> would most likely be a notation for <g\$u>. From this he hypothesizes a form where the 2nd syllable had an /i/,

<sup>6</sup> The Emesal dialect is one of the chief two dialects of Sumerian. On Emesal see Bobrova 1989, Boisson 1989b, Schretter 1990.

<sup>7</sup> Although this conjecture is perfectly sensible (given that the Sumerians are certainly not the inventors of the plough), it is not based on unassailable evidence. On problems related to the spread of agricultural terms in Antiquity, see Blažek & Boisson (1992). The set of words collected in Salonen (1968) to try and prove borrowing by Sumerian from substrates is also characterized by dissimilar vowels.

<sup>8</sup> Edzard as well (p. c.).



not an /u/, something like <g\$uris&><sup>9</sup>. Besides there are cases such as Akkadian <siparru> vs. Sumerian <zabar> "bronze", where it is reasonable to suppose that Sumerian originally had different vowels, so that the earlier form (the one that was borrowed into Akkadian) was \*<zibar><sup>10</sup>. It was only later that vowel harmony came into operation.

Again, in Schretter's list of Emesal words, one finds disyllabic items with dissimilar vowels that are not easily interpreted as compounds, such as "three" <amuš> (Schretter 1990: 154).

In any case there remains a basic problem, which Sumerologists must face squarely: the total vowel harmony attributed to Sumerian by Diakonoff cannot be subsumed under *any of the various types listed in the literature*. Vago (1980) distinguishes between front harmony, labial harmony, and a third type involving the position of the tongue root. See also the brief survey in Smith (1992). It is true that a *tendency* to total vowel harmony is not unknown in a number of languages, notably in the Bantu family, but I know of no case where total vowel harmony would operate as an *absolute rule*<sup>11</sup>. So, for instance, in Tzotzil, a Mayan language of the state of Chiapas, Mexico, the majority of roots are monosyllabic. But there also are disyllabic roots, of the CVCV(C) type. In general, in the case of an adjective, the same vowel appears in both syllables. But neither nouns nor verbs are affected (Haviland 1981: 12-13). Thus vowel harmony in Tzotzil does *not* extend to the whole vocabulary.

<sup>9</sup> But Diakonoff expresses (per lit.) his skepticism with Edzard's reading.

<sup>10</sup> My \*<zibar> (from Lieberman 1977: 70) is to be interpreted phonologically perhaps as something like \*/t<sup>s</sup>ipar/ (see Boisson 1989a). Hayes (1990: 124-125, 228), gives \*<sipar> or \*<sibar>. Of course, one can always argue that this is a loan, a not unreasonable assumption.

<sup>11</sup> Gilbert Puech suggests to me that this could be the case in Maltese, at least for the terms of Arabic origin, but Fernande Krier disagrees.

Another language shows a different phenomenon. "In the Chawchila dialect of Yokuts [...], a word-final vowel is optionally unvoiced. When this takes place, the vowel completely assimilates to the preceding vowel" (Kenstowicz & Kisseberth 1977: 167). Thus one can pronounce *ṭun-k'a* "close the door!" or *ṭun-k'u*; *hiwet-k'a* "walk!" or *hiwet-k'e*. But note that this is optional, and does not apply to the whole word.

Let me give a final observation on this topic. If total vowel harmony seems to be nonexistent on the *phonological* level, it is not unknown on the *phonetic* level. At least this is the situation described for Kalam, a Papuan language of New Guinea (Foley 1986: 51). Kalam is analyzed as lacking phonetic consonant clusters, but as having many on the phonological level. These clusters are automatically broken by a transition vowel which, contrary to the well-known cases in numerous languages, is not simply coloured by the quality of the following vowel, but has to be exactly the same vowel. Thus we have: /kgon/ "garden" = [koŋ gon], /bnep/ "one man only" = [ᵐbenep]. I note this curious possibility, while I doubt very much that Sumerian could offer such phenomena.

If it were so, we would have to develop the following scenario. Let us imagine that the attested form of Sumerian came from an older stage where polysyllables could have different vowels. Let us then imagine that, for instance because of stress, or for some other unknown reason, the vowel of the first syllable was reduced, and became phonologically irrelevant (this development is attested, for instance in Austronesian languages of Indochina, which used to be polysyllabic, but evolved towards more or less strict monosyllabicity). As a purely conjectural illustration, let us take the case of the Sumerian word known under the form <dağal> "be wide". Let us imagine that the early form was an older /tVŋal/ (where /V/ notes a vowel which is unknown, but different from the /a/ which is attested later), which would have gone to /tŋal/, realized with a svarabhakti vowel whose quality would automatically be identical with that of the following vowel. So /tŋal/ would come out as [taŋal], heard as such by Akkadian speakers. This is sheer speculation, hardly more, but such cases of assimilated

anaptyctic (epenthetic) vowels are not unknown. While some languages use a stable svarabhakti to break consonant clusters and optimize the syllable structure, such as [ə], [i] or [u] (examples in Hyman 1975: 146), others insert a svarabhakti vowel that undergoes systematic anticipatory assimilation with the vowel in the next syllable, as is the case in Spanish, or in the Papiamentu creole (Quilis 1970; Holm 1988: 111). Similarly, I have heard an Arabic speaker pronounce French "exprimer" as [ɛksiprime]. This works in the other direction too, since we have both <ag> and <aka> with Vokalzusatz (paragogic vowel), a phenomenon which is perhaps similar to that found in some Atlantic creoles, where the paragogic vowels undergo vowel harmony, e.g. English big, blood, dead > Sranan bigi, brudu, dede (Holm 1988: 111, 124)<sup>12</sup>.

In any case, given the likelihood of an /o/ phoneme, as demonstrated elsewhere (see Lieberman 1977, 1979 for Assyriological arguments; Boisson 1991b for other arguments based on general linguistic constraints), total vowel harmony may be an illusion based on a superficial interpretation of writing. It must be the case that, in a number of dissyllabic words, instead of having an /u/ repeated, we may have /u/ and /o/, or /o/ and /u/. This situation could place Sumerian in the company of well-known cases of languages that have a vowel harmony of the Yawelmani type, where bases "contain two vowels belonging to the same vowel series", as for instance /ɔ·/ then /u/ (Newman 1946: 226). Examples of such reinterpreted words in Sumerian are <domu> "son" (instead of <dumu>), <odu>

12 Incidentally, it is tempting to view Sumerian as a hybridized language: "In any case the languages of ancient empires from China to Sumer expanded along with their military, commercial, and cultural influence, and it is quite likely that this happened via pidginized varieties, although no known records of such speech remain" (Holm 1988: 13-14). But if Sumerian is derived from a creole, then it must have moved a long way from that stage, because, for instance, the intricate system of verbal prefixes is not at all reminiscent of creole structures.

"day" (instead of <ud(u)>), <odug> "spirit" (instead of <ú-du-ug>), <olud> "cup" (instead of <ú-lu-ud>)<sup>13</sup>.

To sum up our conclusions on vowel harmony: (1) There is no known language with total vowel harmony. (2) In Sumerian total vowel harmony is incompatible with the likely existence of an /o/ phoneme. (3) In Sumerian there are earlier forms, or forms in Emesal, without total vowel harmony. Sumerian may have known a tendency to vowel harmony (probably of height<sup>14</sup>) in its earlier stages (although it may have been totally absent in pre-Sumerian), and this tendency gradually became more pronounced in later stages, but a systematic operation of total vowel harmony is ruled out.

Finally, one should not exclude the possibility that, due to the distortions imposed by the writing system, apparently disyllabic words with total vowel harmony are in fact monosyllabic words with consonant clusters.

<sup>13</sup> First example in Hayes (1990: 18); other examples in Lieberman (1977; 1979: 26).

<sup>14</sup> This is the picture that seems to emerge from the studies of prefixes by Poebel and by Kramer, where, after Lieberman's reinterpretation, we might say that the non-high vowel /e/ in prefixes would go with the non-high vowels /a/, /e/, /o/ in the root, while the high vowel /i/ in the prefixes would go with the high vowels /i/, /u/ in the root.

## References

Blažek, Václav and Claude Boisson,

- 1992 "The diffusion of agricultural terms from Mesopotamia." *Archív Orientální* 60: 16-37.

Bobrova, Larisa V.

- 1989 "Ženskij dialekt v šumerskix tekstax svjaščennogo braka." In: *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka. Materialy k diskussijam na Meždunarodnoj konferencii (Moskva, 29 maja - 2 ijunja 1989 g.). Čast' 3.*:16-18. Moskva: Institut vostokovedenija Akademija Nauk SSSR.

Boisson, Claude

- 1989a "Contraintes typologiques sur le système phonologique du sumérien." *Bulletin de la Société de Linguistique de Paris* 84: 201-233.
- 1989b "Emesal: la langue des femmes à Sumer." *La différence sexuelle dans le langage (Actes du Colloque ADEC-Université Paris III, Décembre 1988), Contrastes* (octobre 1989): 79-90.
- 1989c "Lettre aux participants de l'Atelier sur l'indo-européen du Congrès de Pavie, 22 juillet 1985." In: *The New Sound of Indo-European: Essays in Phonological Reconstruction*, ed. Theo Vennemann. Pp. 259-263. Berlin / New York: Mouton de Gruyter.
- 1991a "Note typologique sur les occlusives en étrusque." *Studi Etruschi*, vol. LVI, 1989-1990 (serie III): 175-187.
- 1991b "Topics in Sumerian phonology." Ms.
- 1994 "Conséquences phonétiques de certaines hypothèses de déchiffrement du carien." In: *La decifrazione del cario: Atti del 1° Simposio Internazionale, Roma, 3-4 maggio 1993*, eds. M. E. Giannotta; R. Gusmani; L. Innocente; D. Marcozzi; M. Salvini; M. Sinatra; P. Vannicelli. Pp. 207-232. Roma: Istituto per gli Studi Micenei ed Egeo-Anatolici. Consiglio Nazionale delle Ricerche.

Chadwick, John

- 1987 *Linear B and Related Scripts*. London: British Museum

## Publications.

Civil, Miguel

- 1982 "Studies on Early Dynastic Lexicography I." *Oriens Antiquus* 21: 1-26.

Davidsen-Nielsen, Niels

- 1974 "Syllabification in English words with medial *sp, st, sk*." *Journal of Phonetics* 2: 15-45.

Diakonoff, Igor M.

- 1983 "Some reflections on numerals in Sumerian: towards a history of mathematical speculation." *Journal of the American Oriental Society* 103, 1: 83-93.

Doria, Mario

- 1968 "Les graphies mycéniennes pour s + occlusive initial de mot et de syllabe." In: *Studia Mycenaea*, ed. Antonín Bartoněk. Pp. 59-64. Brno: Universita J. E. Purkině.

Durand, Jean Marie

- 1977 "Diffusion et pratiques des écritures cunéiformes au proche-Orient ancien." In: *L'Espace et la lettre. Ecritures Typographies*. Cahiers Jussieu n° 3, Université Paris 7. Collection 10/18. Pp. 13-59. Paris: Union Générale d'Editions.

Ewen, Colin J.

- 1982 "The internal structure of complex segments." In: *The Structure of Phonological Representations (Part II)*, eds. Harry Van der Hulst & Norval Smith. Pp. 27-67. Dordrecht: Foris.

Falkenstein, Adam

- 1959a *Das Sumerische*. Handbuch der Orientalistik, 1, 2, 1-2, 1. Leiden: E. J. Brill.
- 1959b "Untersuchungen zur sumerischen Grammatik, 5. Zum Akzent des Sumerischen." *Zeitschrift für Assyriologie und vorderasiatische Archaeologie* 19 (53): 97-105.

- Foley, William A.  
1986 *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- Fudge, E. C.  
1969 Syllables. *Journal of Linguistics* 5: 193-320.
- Greenberg, Joseph H.  
1978 "Some generalizations concerning initial and final consonant clusters." In: *Universals of Human Language, volume 2: Phonology*, eds. Joseph Greenberg et alii. Pp. 243-279. Stanford: Stanford University Press.
- Hagège, Claude  
1982 *La Structure des langues*. Paris: Presses Universitaires de France.
- Hagège, Claude and André Haudricourt  
1978 *La Phonologie panchronique*. Paris: Presses Universitaires de France.
- Haviland, John Beard  
1981 *Sk'op sotz'leb: el tzotzil de San Lorenzo Zinacantan*. México: Centro de Estudios mayas, Universidad nacional autónoma de México.
- Hayes, John L.  
1990 *A Manual of Sumerian Grammar and Texts*. ARTANES 5. Malibu: Undena Publications.
- Holm, John  
1988 *Pidgins and Creoles, Volume 1: Theory and Structure*. Cambridge: Cambridge University Press.
- Hyman, Larry M.  
1975 *Phonology, Theory and Analysis*. New York: Holt, Rinehart and Winston.
- Jestin, Raymond  
1943, 1946, 1954 *Le verbe sumérien*. 3 volumes: I.

*Déterminations verbales et infixes, II. Préfixes, particules verbales et noms verbaux, III. Complément.* Etudes orientales publiées par l'Institut français d'archéologie de Stamboul, VII, IX, XIII. Paris: E. de Boccard.

- 1951 *Abrégé de grammaire sumérienne.* Paris: Librairie orientaliste Paul Geuthner.

Kenstowicz, Michael and Charles Kisseberth

- 1977 *Topics in Phonological Theory.* New York: Academic Press.

Kramer, Samuel Noah

- 1936 *The Sumerian Prefix Forms be- and bi- in the Time of the Earlier Princes of Lagaš.* Chicago: Assyriological Studies 8.

Krecher, Joachim

- 1969 "Verschlusslaute und Betonung im Sumerischen." In: *lišān mithurti. Festschrift Wolfram von Soden.* Pp. 157-197. *Alter Orient und Altes Testament* 1.

Laroche, Emmanuel

- 1960 *Les Hiéroglyphes hittites, Première partie: L'écriture.* Paris: Editions du CNRS.

Lejeune, Michel

- 1966 "Doublets et complexes." In: *Proceedings of the Cambridge Colloquium on Mycenaean Studies*, eds. L. R. Palmer & John Chadwick. Pp. 135-149. Cambridge: Cambridge University Press.

Lieberman, Stephen J.

- 1977 *The Sumerian Loanwords in Old Babylonian Akkadian. Vol. I: Prolegomena and Evidence.* Missoula: Harvard Semitic Series 22.
- 1979 "The phoneme /o/ in Sumerian." In: *Studies in Honor of Tom B. Jones*, eds. Marvin A. Powell & Ronald H. Sack. Pp. 21-28. *Alter Orient und Altes Testament* 203. Kevelaer/Neukirchen, Vluyn: Verlag Butzon & Bercker, Neukirchener Verlag.



Lupaş, Liana

- 1972 *Phonologie du grec attique*. Paris / The Hague: Mouton.

Newman, Stanley

- 1946 "The Yawelmani dialect of Yokuts." In: *Linguistic Structures of Native America*, eds. Harry Hoijer & alii. Pp. 222-248. New York: Viking Fund Publications in Anthropology, Number 6.

Palmer, L. R.

- 1963 *The Interpretation of Mycenaean Greek Texts*. Oxford: Clarendon Press.  
1980 *The Greek Language*. London: Faber & Faber.

Poebel, Arno

- 1931 *The Sumerian Prefix Forms e- and i- in the Time of the Earlier Princes of Lagas*. Assyriological Studies 2, Chicago.

Quilis, Antonio

- 1970 "El elemento esvarabático en los grupos [pr, br, tr]." In: *Mélanges offerts à Georges Straka*. Pp. 99- 104. Strasbourg: Société de linguistique romane.

Salonen, Arma

- 1968 "Zum Aufbau der Substrate im Sumerischen." *Studia Orientalia edidit Societas orientalis fennica*: 37, 3: 1-12.

Schretter, Manfred K.

- 1990 *Emesal-Studien, Sprach- und Literaturgeschichtliche Untersuchungen zur sogenannten Frauensprache des Sumerischen*. Innsbrucker Beiträge zur Kulturwissenschaft, Sonderheft 69. Innsbruck.

Smith, Norval S. H.

- 1992 "Vowel harmony and consonant harmony." In: *International Encyclopedia of Linguistics*, vol. 4, ed. William Bright. Pp. 232-234. New York / Oxford: Oxford University Press.

Sollberger, Edmond

1950 "Etudes de linguistique sumérienne." *Cahiers Ferdinand de Saussure* 9: 51-88.

1962 "Graeco-Babyloniaca." *Iraq* 24, 1: 63-72.

Stephens, Laurence and John S. Justeson

1978 "Reconstructing "Minoan" phonology: the approach from universals of language and universals of writing systems." *Transactions of the American Philological Association* 108: 271-284.

Thomsen, Marie-Louise

1984 *The Sumerian Language, An Introduction to Its History and Grammatical Structure*. Mesopotamia, Copenhagen Studies in Assyriology, Volume 10. Copenhagen: Akademisk Forlag.

Vago, Robert M., ed.

1980 *Issues in Vowel Harmony*. Amsterdam: John Benjamins.

Ventris, Michael and John Chadwick

1973 *Documents in Mycenaean Greek*. Second edition. Cambridge: Cambridge University Press.

Walker, Christopher B. F.

1987 *Cuneiform*. London: British Museum Publications.

# The Myth of the Primordial Click

J. C. Catford  
*The University of Michigan*

## INTRODUCTION

In the early part of the twentieth century, a number of scholars expressed the belief that clicks and other “exotic” sounds, particularly all of those with a nonpulmonic airstream, must be relics of a very early stage in the evolution of human speech. Thus, J. van Ginneken (1911) suggested that clicks might be the antique form of many consonants:

Ich glaube, dass wir Westermann’s schon versprochene vergleichende Sudangrammatik abwarten müssen, um zu sehen, ob wir hier nicht den ältesten Lautwert bewahrt finden für all eigenartig gemischten Konsonanten wie *kp*, *gb* usw. in West- und Zentralafrika und die Laute mit Kehlkopfverschluss in Ost-Afrika. Auch in den indogermanischen und semitischen Sprachen finden wir einerseits labiovelare, labiodentale  $\pi\tau$ -  $\beta\delta$ - usw. und so-genannte *ks*-Laute und andererseits auch Laute mit Kehlkopfverschluss, deren Zusammenhang mit den Schnalzlauten zu untersuchen ich mir vorläufig vorbehalten möchte. (p. 347)

Two years later, Sir Harry Johnston, in a book advocating the development of a Universal Alphabet, advanced the same idea (Johnston, 1913). His chauvinistic approach is somewhat comical, if not actually distasteful, to the modern reader. For example, in discussing problems of transcribing exotic languages he says:

Some of the nastiest to tackle are Armenian, the isolated Lesghian or Caucasian group, the tone-using languages of West-central Africa, Bushman, and a good many Amerindian dialects . . . For most of the “unreasonable” languages I have proposed a series of diacritical marks . . . to render their queer and peculiar sounds. (p. 10)

And after scathing criticism of those pedants who insist on using and teaching the native alphabets for Asiatic languages instead of using Latin letters he goes on:

Undoubtedly amongst great Imperial measures to be considered and agreed to by the responsible component parts of the British Empire would be the establishment of a Standard Alphabet to be used throughout the Empire from Ireland to India, Canada to

Australia, in which all languages must be spelt in the schools, colleges and educational institutes, in the rendering of geographical names, in all public and government announcements. (p. 15)

It is not surprising, therefore, that in a footnote, he characterizes clicks as:

half-brutish speech-sounds, vestiges of pre-human speech, resembling the vocal utterances of baboons and apes. Clicks might be called "coarse consonants"; and "consonants"—namely the explosive sounds of tongue, uvula and pharynx—probably preceded vowels in the evolution of human speech. (p. 21)

Here we see three of the beliefs that were to be repeatedly advanced by such scholars as van Ginneken and Stopa, namely, that clicks are "vestiges of pre-human speech", that they resemble "utterances of baboons and apes" and that consonants "preceded vowels in the evolution of human speech".

We will have more to say about the views of van Ginneken and Stopa below, but we must note in passing that the great Danish linguist, Otto Jespersen (1922) also referred to the primeval nature of clicks and other exotic sounds, in these words:

In most languages now, only such sounds are used as are produced by expiration, while inbreathed sounds and clicks or suction-stops are not found in connected speech. In some very primitive South African languages, on the other hand, clicks are found as integral parts of words; and Bleek has rendered it probable that in former stages of these languages they were in more extensive use than now. We may perhaps draw the conclusion that primitive languages in general were rich in all kinds of difficult sounds. (p. 419)

At the 3rd International Congress of Phonetic Sciences at Ghent in 1938 several papers were presented dealing with clicks and other exotic sounds. Among these, a paper on "Evolution in Speech Sounds" by the distinguished Indian linguist S. K. Chatterji, put forward a view of clicks resembling that of Johnston, quoted above.

*Loss of clicks* unquestionably forms another landmark in the evolution of speech sounds: . . . The click sounds are probably to be looked upon as belonging in their function (if not in their formation . . .) to the grunts, croaks, squeaks and screeches and other "non-phonetic" sounds with which the speech of man started from the anthropoid ape stage. (Chatterji, 1939, p. 340)

He goes on to include implosives among "primeval" sounds, in these words:

The presence of a few implosives in (civilised) human speech at the present day is a survival of what may be described as "pre-language" or as the equivalent of speech in primitive man: a wreckage from a richer series, which became merely symbolic in language at large, and which probably at one time were the most easily available phonetic elements when language was forming—to be later on substituted by explosives and other sounds. (p. 341)

It is particularly surprising that Chatterji should have expressed this opinion of implosives, since he must have been perfectly well aware of the relatively recent origin of implosives, and of other exotic sounds, in Indian languages, having published an article on this topic only seven years earlier (Chatterji, 1931).

#### J.VAN GINNEKEN

Van Ginneken contributed a paper entitled "Les clics, les consonnes et les voyelles dans l'histoire de l'humanité," in which he expressed in succinct form views that he stated at greater length in his 1938 *Contribution à la grammaire comparée des langues du Caucase*.

Un clic est un mouvement de succion. Les succions nous sont innées. Chaque enfant commence à produire ces mouvements utiles le premier jour après sa naissance sans que personne ne le lui ait appris . . . Or en l'absence de la mère, chaque enfant normal dans le deuxième ou troisième mois de son existence commence à tenir des repas en imagination. Seulement l'air inspiré y prend la place du lait maternel; et cette succion d'air est perceptible à l'oreille du nouveau-né, qui s'en amuse à merveille.

Comme phonèmes lexicaux les clics inspiratoires sont devenus rares dans les langues d'aujourd'hui. C'est que peu à peu ils ont été remplacés par les groupes de consonnes expiratoires. Dans l'Afrique du sud la plupart des mots commence encore par un clic et la première partie du clic est encore inspiratoire, mais la deuxième partie est devenue déjà expiratoire. (van Ginneken, 1939, p.322)

Van Ginneken goes on to claim that the same thing happens in the Caucasus, where, in Mingrelian the first part of lateral clicks still exists as a palatal inspiration, immediately followed by a laryngeal and an expiratory lateral:

cql. Needless to say, this is quite untrue. There are no "sons inspiratoires" as phonemic norms in any Caucasian languages and no clicks.<sup>1</sup>

Van Ginneken continues:

Je puis prouver avec plus ou moins de sureté la même chose pour les langues négro-africaines avec leur labiovélares, pour les langues hamito-sémitiques avec leur racines trilitères et leurs consonnes emphatiques, pour les langues ouralo-altaïques avec leurs clics latéraux gardés en ostyak . . . pour les langues austronésiennes avec huit clics latéraux dans leur langue commune, dont au moins deux existent encore dans le dia-lecte de Kambara à l'île de Soumba; et pour assez bien des familles de langues dans les deux Amériques qui ont gardé aussi jusqu'à ce jour plusieurs clics latéraux. (p. 323)

Van Ginneken seems to have been obsessed with the idea that lateral obstruents (i.e., lateral affricates and fricatives) are clicks, or at least must be recently derived from clicks. Indeed, in van Ginneken (1938, p. 7) he tells us that, at the first International Congress of Linguists at the Hague "MM. Jacovlev et Dirr m'en fait entendre plusieurs fois les différentes consonnes latérales, *et depuis j'ai l'idée fixe que les affriquées du moins sont des clics à succion injective* [italics added]. . ." Jakovlev, certainly, was an excellent phonetician (I am not so sure about Dirr) and presumably pronounced the sounds accurately. Van Ginneken simply misheard, and misanalysed them.

His general belief in the existence of ingressive sounds in Caucasian was also supported, as he says, by an observation of Vogt (1936, p. 11), who says, referring to the Georgian sequence tq: "En effet les tracés graphiques présentent parfois dans le groupe tq le q comme un son à succion". Vogt here was no doubt echoing the words of Selmer (1935, p. 49) ". . . der Character der *occlusive postvélaire* (q) mir noch immer dunkel und rätselhaft geblieben ist" followed by the tentative suggestion that it may be an "implosiver Suctions-laut?"

It is also not improbable that van Ginneken was influenced (and misled) as well by the kymograph tracings of "sons inspiratoires" in L'Abbé Rousselot's *Principes de Phonétique Expérimentale* (1897-1901, T.1, pp. 489-494). These include short utterances supposedly in Circassian and Georgian (pronounced, as the Abbé tells us by H. Adjarian, author of *Classification des Dialectes Arméniens*, and two other Armenians!). The Circassian tracing purports to represent the Circassian word for *bird* which is described by Rousselot as "un mot formé de la syllabe *dzi*, puis d'une sorte de hoquet qui se termine par la

<sup>1</sup> There is one authenticated exception to this! Kibrik and Kodzasov (1990, p. 338) mention that in the Burshag dialect of Agul the verb 'to kiss'—*pac / aq'as*—has a bilabial click as its initial consonant; but this click is obviously a mere gestural or onomatopoeic exception, and not a regular phoneme of Burshag.

voyelle *u*: en réalité deux syllabes, l'une expiratoire, l'autre inspiratoire." This strange hiccupping word does not resemble any Circassian form that I can identify, unless it be *ts'əkʷ* 'little' in the compound *bzəw-ts'əkʷ* 'little bird'. In an exaggerated imitation the abrupt mouth-opening with the glottis still closed for the ejective *kʷ* could well induce the very noticeable influx of air seen in the kymogram.

All these examples, from Selmer and Rousselot, certainly show ingressive airflow into the mouth; but this is due simply to the lowering of intra-oral pressure by rapid mouth opening,—a not uncommon purely articulatory effect, with nothing to do with the initiation (air-stream mechanism) of the sounds. Indeed on page 494 Rousselot presents tracings of the syllables *na* and *la*, pronounced by a Russian, which show a sudden brief influx of air at the transition from the consonant to the vowel. This again, is purely articulatory, resulting from the sudden increase in mouth-volume due to the rapid transition from a strongly velarized *n* or *l* to an open vowel. It is easy to replicate this effect, which I have done several times when airflow-recording equipment was available.

In the 1939 article, van Ginneken summarizes his views (pp. 324–325) as follows:

1. All consonants are issued from *inspiratory clicks* or *injectives*.
2. Clicks are transposed first into consonantal groups, half inspiratory, half expiratory, finally becoming wholly expiratory.
3. Consonant groups are thus more primitive than all simple consonants.
4. *Affricates* seem to be early consonants, only split later into *stops* and *fricatives*.
5. *Sonants* derive from consonant groups with their origin in the lateral clicks; the group *muta cum liquida* is thus more primitive than the simple liquids.
6. All vowels originate from more or less open and closed consonants, functioning for some time as "semblants de voyelles" (he has in mind the 3-vowel, ə-e-a system of the Circassian (Adygian) languages), etc.

## ROMAN STOPA

I turn now to Roman Stopa, who, since the publication of his book *Die Schnalze* in 1935, (cited in van Ginneken, 1938, p. 6) continued to champion the idea that the clicks that he had observed in Bushman were relics of primeval speech in his 1939 paper to the Phonetic Congress, as well as further publications in 1962, 1965, 1972, 1979, and 1983.

The 1939 Congress paper *Die Schnalze* begins (Part I) with a description of the nature of clicks. Part II describes the types and functions of clicks in the sound system of the Nama dialect of Hottentot. It is clear from this, and also from the excellent descriptions and diagrams in some of Stopa's later works, that by clicks he means precisely what is meant by the term in all modern phonetic

publications, namely: velaric ingressive stops and affricates. This should be borne in mind when we discuss the question of clicks in the vocalizations of nonhuman primates. Then on page 331, Part III he lists the distribution of clicks in languages:

1. Primary click languages: Bushman, Hottentot, Sandawi, Hadzapi (also known as Wakindiga)

2. Secondary click languages: Southern Bantu (Zulu-Kaffir, Sotho. Swazi); these presumably have a click-language substratum.

3. Languages of Pygmies: As relics, clicks are found among almost all peoples as interjectional forms:

4. That clicks here are to be interpreted as a remnant of older speech periods is suggested *inter alia* by three facts:

- a. They are used in children's speech

- b. They occur in the speech of the deaf and dumb, i.e. in the initial attempts of deaf-mutes to imitate vocal speech.

- c. Many animals produce clicks, e.g. Apes. People also often address animals with clicks, wherein lies a latent supposition that the expressive value of clicks is intelligible to animals . . . Accordingly, the clicks appear to belong to the oldest speech manifestations of mankind. Probably many physical, psychical, and perhaps also cultural circumstances have contributed to the fact that these sounds were retained until today in the languages of the Bushmen and Hottentots.

In his 1962 article, "Bushman as a language of primitive type," Stopa enlarges on the claim that clicks are found in infant "speech" stating that "we find inspiratory and injective, clicking and ejective (disjunctive) sounds as early as the first year" (p. 190). Incidentally, in his examination of Stopa's claim that Bushman is a language of primitive type, Pesot (1983) was apparently misled into believing that Bushman has "ejectives that have not been detected elsewhere except in the vocalizations of monkeys, infants, deaf-mutes and tracheotomized persons . . ." (p. 517). This, of course, is incorrect, since ejectives are found as regular phonemic norms in about 20% of the world's languages, but are seldom, if ever, reported in the speech of monkeys and infants, though they are certainly used by tracheotomized persons, and possibly by deaf-mutes.

With respect to the existence of clicks in the early utterances of human infants—a claim made strongly by both van Ginneken, and Stopa—it is interesting to note that at that same 1938 Phonetic Congress P.de V Pienaar (1939) in a very well-informed survey of the nature and distribution of clicks says:

With regard to the theory that clicks are the most primitive sounds of mankind, I have recorded this fact that the Bantu and Hottentot children, when they acquire the language from their parents, at first have great difficulty with the click sounds. (p.



353)

There follow some examples of substitutions for clicks made by children in Zulu, and as given by a Koranna (Hottentot) informant telling an animal story in which the animals were supposed to speak like children.

In fact, the sucking activities and the resultant suction sounds of infants, that van Ginneken made so much of, are quite non-social and non-communicative, and they are, in any case far outnumbered by the vowel-like pulmonic egressive cries which infants produce from birth, and which quickly acquire communicative functions.

In his 1972 book, *Structure of Bushman and its Traces in Indo-European*, Stopa discusses the expressive use of clicks by nonhuman primates citing (p. 11) Andrew (1964). He takes up this theme again in his 1979 *Clicks: Their Form, Function . . .* referring to the same article by Andrew, of which more below.

In addition, in the latter work he refers (p. 15) to Kenneth Hale's (1973, p.443-444) description of Damin, initiation language of the Lardil people of Mornington Island, Australia. Damin has four velaric ingressive (clicks) as well as an ejective (glottalic egressive) k' a pulmonic ingressive voiceless lateral, and a velaric egressive bilabial stop. Stopa likens the Damin clicks to those of Bushman and cites Bushman parallels to the Damin words with initial clicks cited by Hale. He then concludes: "At any rate we may assume that the initiation jargon Damin had its source in a clicking language, and that Australian languages were clicking in their remotest history" (p. 16).

However, no other Australian language uses clicks, and there is no reason to suppose that these languages were formerly click-using. As I have pointed out elsewhere (Catford, 1974, p. 28, 1977, p. 65), the four nonpulmonic-egressive air-stream mechanisms of Damin are unique in Australia, and two of them (pulmonic ingressive and velaric egressive) are unique in the entire world as the air-stream mechanisms of regular segments in words. The extraordinary and unique exuberance of air-stream mechanisms in Damin tempts one to hypothesize that the Damin sound-system is a deliberately invented one. Damin does not provide convincing evidence for the existence of clicks in proto- or early Australian, or for clicks as being typical sounds of "primitive" languages.

## ALLEGED USE OF CLICKS BY NONHUMAN PRIMATES

As we have seen, Johnston, Chatterji, and above all Stopa, have cited the existence of clicks in the utterances of nonhuman primates as evidence for the probable existence—indeed the exclusive use—of clicks in the earliest stages of human speech. But what does this evidence amount to?

Stopa (1972, p. 11, 1979, pp. 36, 41, 56) refers to R. J. Andrew's (1964) article "Displays of the primates". This article deals with the vocalizations of a number of small primates which are about as remote from *homo sapiens* as any primates can be, namely, the little shrew *sorex palustris* and types of loris and

lemur. Andrew presents spectrograms, some of which show what are called clicks. These (pp. 285, 292–293) are not very clear and have a very small time scale, but some, at least, look more like brief squeaks or other voiced sounds than anything resembling Bushman clicks!

One must always remember, when looking at such records, that what a biologist calls a click is defined purely acoustically, and its production may be totally different from that of the human velaric ingressive sounds labelled clicks by phoneticians. Indeed, in an article on bioacoustic terminology Broughton (1963) is quite explicit about this:

Thus, the essence of the click is its short duration in time and its discreteness . . . it is perhaps desirable to emphasize that no more than a pulse (of which it is a special case) should a click be defined in terms of its generation or reception. (p. 12)

There are some more relevant studies of nonhuman primate vocalizations. One of these is Marler (1969) which enumerates wild chimpanzee vocalizations described as *grunts*, *rough grunts*, *barks*, *screams*, *hoots*. The only sound approximating an (acoustic) click is a *squeak*, which is described as “a shortened sound with essentially the same spectral structure as a scream . . .” The spectrogram of five consecutive squeaks indeed shows sound bursts ranging from about 2 cs. to 10 cs. duration (mean 6 cs.) with almost exactly the same formant structure as screams, of which spectrograms are also given. But these, of course, have nothing in common with Bushman or other human velaric clicks.

Andrew (1976) points out that virtually no attention has been paid to calls very like those of man in their tonal structure and resonances (“humanoid grunts”) and the article goes on to illustrate, with spectrograms, many such types of vocalization by baboons and other nonhuman primates. Of particular interest to us is the fact that although the article refers to 16 or 17 different kinds of humanoid vocalization (namely: *screams*, *barks*, *shrieks*, *squeals*, *trills*, *long moans*, *grunts* (*woof & waa*), *raugh*, *coo*, *woo*, “*wahoo*” *bark*, *girning*, *screeches*, *breathy calls*, *low panting calls*, and *whoops*) the only click referred to is the occasional sound of the teeth clicking together (“tooth click”). Not only is this bidental percussive click completely outnumbered by the various types of voiced sound, but in its production it clearly has no relationship whatsoever to the velaric ingressive clicks of Bushman, etc.

And in case anyone should think that van Ginneken’s bizarre hypothesis of the clicking origin of most Caucasian consonants is supported by this bidental percussive—as a possible source of the unique bidental fricative of the Shapsug dialect of Adyghe—I hasten to point out that we know perfectly well that this unusual fricative is simply a reflex of Proto-Circassian x. Relaxation of the

velar stricture for *x* allowed the airstream to have sufficient volume velocity to generate turbulence as it passes by the more or less clenched teeth.<sup>2</sup>

One other article on primate calls is Newman (1992) where we read:

A comparative analysis of the isolation calls [distress sounds of infants separated from caregivers] of more than 20 primate species (representing Prosimians, Old World and New World monkeys, Great Apes and humans) indicates that, with a few notable exceptions, the overall acoustic structure of the isolation call is always the same, namely, a tonal or voiced sound with little in the way of noise or acoustic transients, or a repeated series of sounds with these basic acoustic characteristics. p. 304)

Newman's Table 2 (p. 305) lists 44 examples of isolation call characteristics. Only three of these are described as simply "clicks" two others being "clicks or squeals" and "click, zek", and these few clicks are listed only for the smallest primate species, as remote from Great Apes and humans as possible. These, we may be virtually certain, are merely "acoustic clicks" with no resemblance to the velaric ingressive clicks of human speech.

So, there seems to be no justification for linking Bushman and Hottentot velaric ingressive clicks with the sounds produced by nonhuman primates.

## ORIGINS OF "EXOTIC" SOUNDS: LATERALS

As I suggested in Catford (1974, p. 21), it seems probable that this once popular view that various exotic sounds, such as lateral obstruents, ejectives, and implosives, must either themselves be primeval, or else be derived from the primordial click, was due to failure to understand how any of these "exotica" could have resulted from perfectly comprehensible mutations of more "normal" sounds. Now, with more information on out-of-the-way languages, and more sophisticated insight into certain aspects of phonetics—particularly the aerodynamics of speech production—we need not resort to fantastic explanations.

Van Ginneken, as I mentioned above, was obsessed with the idea that the lateral fricatives and affricates of North Caucasian languages must be derived from clicks—and not only those, for he refers (1938, p. 6) to the voiceless lateral fricative *ɬ* of Welsh and (1939, p. 323) that of Ostyak (Khantiy), not to mention "les langues austronésiennes avec huit clics latéraux dans leur langue commune" (What can those "eight lateral clicks" be?) and of course "assez bien de familles de langues dans les deux Amériques qui ont gardé aussi jusqu'à ce jour plusieurs clics latéraux."

<sup>2</sup> Not a reflex of *xʷ*, as incorrectly stated by Colarusso (1988, p. 57). The Shapsug reflex of *xʷ* is *f* as in other dialects of Adyghe.

The origins of the North Caucasian and American lateral obstruents, which include both pulmonic and glottalic (ejective) lateral fricatives and affricates (but no "clics latéraux" of course) is unknown. They are assumed to have existed in the North Caucasian proto-language, and presumably in one or more American proto-languages as well. If we accept Nikolaev and Starostin's Dene-Caucasian hypothesis they no doubt also existed in Proto-Dene-Caucasian.

But although the remote origin of Dene-Caucasian lateral obstruents may be unknown, that is no reason to assume that they are necessarily derived from clicks. There are numerous examples of lateral fricatives and affricates whose origin *is* known; and none of these is derived from a click.

In some cases a voiceless lateral fricative has apparently arisen from the simple devoicing of initial (or final) *l*, for instance, Welsh *l̥* in *llau* "lice" (cf. OHG *lūs*), Eng. *louse*, W. *llesg* "weak" (cf. ON *lōskr* "dull") and possibly this is also true of the development of voiceless *l* in the Dravidian languages Brahui and Toda.

More commonly, however, voiceless lateral obstruents have arisen as a result of phonatory assimilation from a contiguous voiceless sound, thus, for example, W. *llau* "host" (cf. OIr. *sluag*, O.Slavic *sluga* "servant").

In other cases the assimilatory genesis of a lateral obstruent is more complex. Thus, in the Bhalesi dialect of Bhadarwahi in Jammu and Kashmir *br* and *dr* have become some form of *dl*, and *pr* and *tr* have become *t̪* (or more precisely *t̪* with a retroflex stop) (Bailey, 1908; Varma, 1948). Two hundred miles northwest of Bhalesi, in a number of Central Dardic languages, the sequence *tr* has lost its stop altogether, yielding a simple voiceless lateral fricative e.g., *te* "three" (<*tre*), *puṭ* "son" (<*putr*-) etc., (Edelman, 1983). A quite similar change of *tr* (and sometimes *kr*) to *t̪* has occurred in the South Thai dialects of Songkhla and Ranot (Brown, 1965).

In other cases *t̪* has arisen directly as a mutation of a voiceless sibilant, *s* or *ʃ*—the phenomenon known to speech therapists as "lateral sigmatism". This kind of lateral sigmatism apparently became endemic in the Sze-Yap ("four towns") dialect of the Canton River basin of South China, where *t̪* has replaced ancient Chinese *\*s* and *\*ʃ*, as in:

Ancient Chinese	<i>*-sam</i>	Sze-Yap	<i>-t̪am</i>
"three"			
Ancient Chinese	<i>*\ʃiem</i>	Sze-Yap	<i>-t̪im</i>
"deep"			

There is no need to assume (with van Ginneken) that obstruent laterals are derived from clicks.

It may be worth pointing out that this opinion is not contradicted by a forthcoming article by Traill and Vossen (1997, in press) which cites examples of the replacement of clicks by lateral obstruents in some Khoisan languages.

The situation there is not the same, however, since it concerns attested cases of click replacement. The Van Ginneken cases concern attested lateral obstruents which are gratuitously assumed to be derived from clicks.

## IMPLOSIVES

As for implosives (glottalic ingressive stops, usually voiced), which Jespersen, Chatterji and van Ginneken all regard as either primeval or derived from clicks, there is one case, at least, where their origin is well known. This is the case of the Sindhi and Lahnda (Hindki or Multani dialect) implosives, *ɓ* *ɗ* etc. Turner (1927) showed that, initially, these sounds correspond to Skt. *g* -, *j* - (*dy*-), *ḍ*-, *b*- (*dv*-) and intervocalically to Skt consonant groups that became -*gg*-, -*jj*-, -*ḍḍ*-, (-*ḍḍ*-), -*bb*-, (-*vv*-) in Pkt. The mechanism of this change is not mysterious. The mediating factor is the downward larynx movement or pharynx expansion that ensures the voicing of a voiced stop. A slight exaggeration of this, together with diminution of the pulmonic activity, leads to the implosive mutation.

There are voiced implosives in a good many other languages (about 10% of all languages) where we do not necessarily know their origin. When we plot the locations of languages using implosives on a map of the world, (see Map 1, p. 70)<sup>3</sup> we find that, with few exceptions, they occur in Africa and South East Asia, concentrated in a band lying between the Equator and the Tropic of Cancer. Their location in Africa is largely coextensive with an area where doubly articulated plosives of the labial-velar (*gb*) type are commonly found, and in some cases, at least, these may be the source of implosives. Where we have a coarticulation of this kind, small incidental movements of the articulators can bring about pressure changes in the mouth. If the pressure of the air contained between the labial and velar articulations of *gb* is lowered by a small backward movement of the tongue, there will be a slight influx of air into the mouth when the articulatory closures are released. At this stage we have an incipient implosive (and, incidentally, an incipient click), and if the weakly ingressive character of the labial release is found to differentiate the sound sufficiently from a plain *b*, the implosive may well lose its velar component and be stabilized and institutionalized in the language.

The situation in the South East Asian implosive area is different. Here, double articulations are rare, and in many, or most, cases the voiced implosive is not in contrast with a plain voiced plosive. In this case the implosive seems to have evolved spontaneously, as it were, not as a result of overlapping or double articulation. A tendency to generate implosives in this way is not particularly

<sup>3</sup> The data for the Maps is taken from Ruhlen (1975) and Maddieson (1984) A first draft of those, and many other linguistic maps, was made by my students Ümit Alacahanlı, Jilda Baronyan, Oğuz Baykara, Nafi Yalçın, at Boğaziçi Üniversitesi, Turkey.

surprising since this is a case where we can reasonably say that one sound is easier (requires less muscular effort) to produce than another.

The production of voice requires a pressure drop across the glottis of about 2 to 3 cm H<sub>2</sub>O to keep air flowing and the vocal folds vibrating. The necessary pressure difference can be achieved in either of two ways—by increasing the subglottal pressure by compression of the lungs, or by lowering the supraglottal pressure by slightly lowering the larynx and enlarging the pharynx. To ensure retention of the necessary pressure difference throughout the stop, we might start with a pressure of, say, 6 cm H<sub>2</sub>O. Assuming a subglottal volume of about 2000 cm<sup>3</sup>, in order to reach that pressure for a pulmonic egressive b, we must reduce the subglottal volume by about 12 cm<sup>3</sup>. On the other hand, for a glottalic ingressive β, in order to reach the equivalent pressure (in this case, a negative pressure of -6 cm<sup>3</sup>) we have to enlarge the supraglottal volume (of about 160 cm<sup>3</sup>) by only 0.9 cm<sup>3</sup>. This small pharynx expansion presumably takes considerably less effort than the much larger compression of the thoracic cage required for voiced pulmonic plosives.

It is, perhaps, no accident that these sounds, which require less energy for their production than pulmonic egressive voiced plosives, should cluster thickly in this zone, between the Equator and the Tropic of Cancer, where the climate dictates that the expenditure of unnecessary energy is undesirable.

The whole question of the nature, functions, and origins of glottalic consonants, especially implosives, was discussed at length in the important article by Greenberg (1970).

## EJECTIVES

Ejectives are another class of sound that has been regarded as primeval or as necessarily derived from clicks. Once again, there is some evidence of the genesis of ejectives from "normal" sources.

Chatterji (1931, 1960, pp.112-115) describes the rather complex genesis of what are apparently ejectives (as well as implosives)—though the descriptions of these sounds are not very clear—as mutations of pulmonic egressive p, t, etc. in several Indian languages, particularly Gujarati and East Bengali dialects, and somewhat similar developments have occurred in Western Pahari dialects, for instance, N. Jubbali (Bailey, 1920, pp. xiii, 172). These developments seem to involve the rather surprising change of h (with open glottis) to ʔ (with closed glottis), but with additional complications. The important point, however, is that ejectives (and probably implosives as well) have arisen here simply as mutations of "normal" sounds—not as reflexes of clicks or relics of primeval speech.

Elsewhere ejectives may have arisen in other ways. They are, as we know, endemic in Caucasian languages, all 37 of which have ejectives alongside two, or in some cases, three series of pulmonic egressive obstruents. Two Indo-

European languages of the Caucasus area also have ejectives, namely: Ossetic and Eastern Armenian. Ossetic appears to have acquired its ejectives by adoption, in words borrowed from N.Caucasian. In Armenian, however, the situation is different. Here the ejectives (according to the traditional view) result from mutation of the IE voiced plosives, presumably from interaction between the simultaneous components initiation (air-stream mechanism), phonation and articulation (as in the case of the Sindhi implosives).

It is conceivable that the migration of the Armenian people from near sea level to the high terrain of ancient Armenia in Eastern Anatolia might have contributed to this change. As I have demonstrated elsewhere (Catford, 1974, p. 25), there is a good aerodynamic reason that it is slightly more difficult to produce voiced stops at an altitude of five or six thousand feet above sea level (Lake Van in the heart of ancient Armenia, lies at 5,644 ft.) than at elevations near sea level. If the Armenians did, in fact, have a tendency to produce ejectives in place of pulmonic voiced plosives, because of altitude or for any other reason, their proximity to the ejective-using Kartvelians would have facilitated the adoption of ejectives as phonemic norms by what I have called "sympiotic selection". Of course, if one accepts the "glottalic theory" regarding IE plosives, then the Armenians, unlike all other IE speakers (including the Ossetes) would simply have retained the original IE ejectives—perhaps for similar reasons to those just mentioned.

The idea that altitude might contribute to the occurrence (or prevention) of a sound change has been scoffed at in the past. Nevertheless, since the production of speech sounds is primarily an aerodynamic process, it would not be surprising if high altitude, with the related lowering of ambient air pressure, sometimes played a part in some sound mutations, such as the generation of ejectives. It is a curious and suggestive fact that, if we plot the incidence of ejectives on a map of the world, we find that they cluster most thickly in mountainous areas. We have already mentioned the Caucasus where the majority of the indigenous inhabitants have lived for many thousand years at high altitudes. But the two other regions of the world where ejectives are most common are the mountainous Ethiopian region of Africa, and the mountainous west coast of the Americas, particularly North America. In the rest of Africa and the American continent ejectives of course occur, but they are very thinly spread. (See Map 2, p. 71)

## CLICKS

We must now consider the genesis of clicks. Even if we know little or nothing about the origin of clicks in the Khoisan languages, it is not difficult to understand how clicks can arise as a result of slight mistimings and small movements of overlapping or double articulations. That is to say, adventitious clicks (and implosives) can, and do, easily arise in such circumstances, as we have already seen. Indeed, the genesis of such sounds from overlapping articulations in Shona dialects is mentioned by Doke (1931, pp. 123, 139).

A formerly well-known example is the notorious Breton bilabial click. This is an adventitious momentary (and presumably velaric) ingressive click occurring at the transition from *m* to *k* in *toram kwat* "let's cut wood". A kymogram in Rousselot (p. 493) clearly shows momentary ingressive airflow between the *-m* and the *kw-*. This was referred to by Vendryes (1923, p. 39) and, of course, was seized on by van Ginneken and linked with the Welsh *ll* as "survivances des clics" (1938, p.18). And we have mentioned above other examples of adventitious ingressesives, observed in kymograms by Selmer and Rousselot.

## CONCLUSION

It seems clear, then, that there is no reason at all to assume that clicks and other "exotic" sounds—particularly lateral obstruents, implosives, ejectives—must necessarily be relics of a very ancient stage in the development of human speech, or that clicks themselves are the primordial speech sounds, from which all the rest have been derived. As we have seen, none of the supposed evidence for the primordial status of clicks, presented or implied by van Ginneken, Stopa and others is at all convincing.

Though clicks are very much used in Bushman—a "primitive" language of a "primitive" people, according to Stopa—they simply do not occur in the languages of any other "primitive" peoples. Damin is no evidence for the pre-existence of clicks in Australian languages.

Infants in the first few months of life may, as van Ginneken claims, sometimes indulge in solitary clicking. But very much more frequently they produce sounds of many other types—mostly pulmonic egressive—and rarely, if ever, make any communicative use of clicks. Moreover, as Pienaar (1939) pointed out, when they really begin to acquire language sounds they have difficulty integrating clicks into their speech. In any case, it is doubtful if Haeckel's "biogenetic law" can legitimately be applied to a cultural process like the acquisition of the sounds of one's language.

Nonhuman primates, particularly those most closely related to man, apparently make little or no use of clicks (in the phonetically relevant sense of velaric ingressesives). By far the greatest part of their communication seems to be carried on by means of pulmonically initiated "grunts" "screams" "squeals" etc.

Finally, all of the exotic sounds that have seemed to some to be so strange that they must have had some special, aberrant, fantastic origin can be perfectly well related to all human speech sounds within a framework of phonetic categorization that takes note of the aerodynamics of speech and the relations and interactions between the functional components of speech production—initiation (of an air-stream), phonation, and articulation.

It would, indeed, have been quite extraordinary if early hominids should have chosen to communicate by clicks rather than by the simple pulmonic grunts and squeals that they were undoubtedly able to produce just as their primate ancestors did. And surely, at a later stage of language evolution, when *Homo sapiens'* upper respiratory system was already similar to ours—perhaps



about 400,000 to 300,000 years ago (Laitman, 1983, p. 84)—and presumably something approximating to phonologically and syntactically fully articulated language was developing, it would have been most natural to use pulmonic egressive sounds.

As pointed out in Catford (1974):

The pulmonic initiator provides by far the most copious store of air for speech: speakers normally use about 500 to 1000 cm<sup>3</sup> between inhalations, and this means, at normal volume velocities of 100 to 300 cm<sup>3</sup>/sec., duration of breath-groups of up to 10 seconds. In other words, one can talk for as much as 10 seconds without 'recharging the *pulmonic* initiator', whereas the *glottalic* initiator (using only supraglottal air) must be recharged every second or two at most, while the *velaric* initiator (using a very small quantity of air in the mouth) can produce only momentary sounds. In the second place, of the two directions of initiation, egressive and ingressive, the shape of the vocal cords (which present an inclined surface and hence a nozzle-like channel to airflow from below) renders them much more suitable for the generation of *egressive* voice than *ingressive* voice. It is not surprising, therefore, that pulmonic egressive is the initiation type used for all sounds in very many languages and for most sounds in all languages. (p. 24)

It seems most reasonable, then, to assume precisely the *opposite* of the van Ginneken-Johnston-Jespersen-Chatterji-Stopa primordial click hypothesis—namely that human speech began with pulmonic egressive initiation and simple articulations, and that in the course of development of language, clicks and other exotic sounds have arisen here and there as mutations of the normal pulmonic egressive types of sound.

## REFERENCES

Adjarian, H.

1909 *Classification des dialectes Arméniennes*. Paris: Champion.

Andrew, R. J.

1964 "The displays of the primates." In J. Buettner-Janusch (Ed.), *Evolutionary and genetic biology of primates: Vol. 2* (pp. 284–294). New York: Academic Press.

1976 "Use of formants in the grunts of baboons and other nonhuman primates." In S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds.), *Origins and evolution of language and speech* (pp. 673–693). *Annals of New York Academy of Sciences*, 280, 673–693.

Bailey, T. G.

1908 *The languages of the northern Himalayas*. London: Royal Asiatic Society.

1915 *Linguistic studies from the Himalayas*. London: Royal Asiatic Society.

Broughton, W. B.

1963 "Method in bioacoustic terminology." In R. G. Bushel (Ed.), *Behaviour of animals* (pp. 3–24). Amsterdam/London: Elsevier.

Brown, J. M.

1965 *From ancient Thai to modern dialects*. Bangkok: Social Science Association Press of Thailand.

1985 *From ancient Thai to modern dialects: And other writings on historical Thai linguistics*. Bangkok: White Lotus Co. Ltd.

Catford, J. C.

1974 "'Natural' sound changes: Some questions of directionality in diachronic phonetics." In A. Bruck, R. A. Fox, & M. W. La Galy (Eds.), *Papers from the parasession on Natural Phonology, April 18, 1974* (pp. 21–29). Chicago: Chicago Linguistic Society.

1977 *Fundamental problems in phonetics*. Bloomington/Edinburgh: Indiana University Press/Edinburgh University Press.

Chatterji, S. K.

1931 "Recursives in new Indo-Aryan." *Indian Linguistics*, 1(1), 15–64.

1939 "Evolution in speech sounds." In E. Blancquaert (Ed.), *Proceedings of the 3rd International Congress of Phonetic Sciences* (pp. 332–344). Ghent: Laboratory of Phonetics of the University.

1960 *Indo-Aryan and Hindi*. Calcutta: Firma K. L. Mukhopadhyay.

Colarusso, J.

1988 *The northwest Caucasian languages*. New York: Garland.

Doke, C. M.

1931 *A comparative study in Shona phonetics*. Johannesburg: University of Witwatersrand Press.

Edel'man D. I.

1983 *The Dardic and Nuristani languages*. Moscow: Nauka.

van Ginneken, J.

1911 "Sprachwissenschaftliche Chronik" [Linguistic chronicle], *Anthropos*, 6, 347.

1938 *Contribution à la grammaire des langues du Caucase*. Amsterdam:

N.V. Noord-Jollandsche Uitgevers-Maatschappi.

- 1939 Les clics, les consonnes et les voyelles dans l'histoire de l'humanité. In E. Blancquaert (Ed.), *Proceedings of the 3rd International Congress of Phonetic Sciences* (pp. 332–344). Ghent: Laboratory of Phonetics of the University.

Greenberg, J.

- 1970 "Some generalizations concerning glottalic consonants, especially imlosives." *International Journal of American Linguistics*, 36, 123–145.

Hale, K.

- 1973 "Deep-surface canonical disparities in relation to analysis and change: An Australian example." In T. A. Sebeok (Ed.), *Current trends in linguistics* (pp. 401–458). The Hague: Mouton.

Jespersen, O.

- 1922 *Language: Its nature, development and origin*. London: G. Allen & Unwin Ltd.

Johnston, Sir Harry

- 1913 *Phonetic spelling: A proposed universal alphabet for the rendering of English, French, German and all other forms of speech*. Cambridge: Cambridge University Press.

Kibrik, A. E., & Kodzasov, S. V.

- 1990 *Sopostavit'elnoe izuč'enie dagestanskix jazykov: imja, fonetika* [A comparative study of Dagestanian languages: Noun, phonetics]. Moscow: Izdat. Moskovskogo universiteta.

Laitman, J. T.

- 1983 "The evolution of the hominid upper respiratory system and implications for the origin of speech." In E. de Grolier (Ed.), *Glossogenetics: The origin and evolution of language: Proceedings of the International Transdisciplinary Symposium on Glossogenetics* (pp. 63–90). Paris: Harwood Academic Publishers.

Maddieson, I.

- 1984 *Patterns of sounds*. Cambridge: Cambridge University Press.

Marler, P.

- 1969 "Vocalizations of wild chimpanzees." In *Proceedings of the 2nd International Congress on Primates: Vol. 1* (pp. 94–100). Basel, New York: Karger.

Newman, J. D.

- 1992 "The primate isolation call and the evolution and physiological control of human speech." In J. Wind et al. eds., *Language origin: A multidisciplinary approach*: 301–321. Dordrecht: Kluwer Academic Publishers.

Pesot, J.

- 1983 "On the direct study of the phylogeny of language." In E. de GrolierEd., *Glossogenetics: The origin and evolution of language: Proceedings of the International Transdisciplinary Symposium on Glossogenetics*: 513–525). Paris: Harwood Academic Publishers.

Pienaar, P. de V.

- 1939 "Click formation and distribution." In E. Blancquaert, ed., *Proceedings of the 3rd International Congress of Phonetic Sciences* (pp. 332–344). Ghent: Laboratory of Phonetics of the University.

Rousselot, P. J.

- 1897–1901 *Principes de phonétique expérimentale: Vol. I*. Paris: H. Welter.

Ruhlen, M.

- 1976 *A guide to the languages of the world*. Stanford, California: The Author.

Selmer, E. W.

- 1935 *Georgische experimentalstudien*. Oslo: I Kommissjon Hos Jacob Dybwad.

Stopa, R.

- 1939 "Die Schnalzlaute." In E. Blancquaert, ed., *Proceedings of the 3rd International Congress of Phonetic Sciences* (pp. 332–344). Ghent: Laboratory of Phonetics of the University.
- 1962 "Bushman as a language of primitive type." *Folia Orientalia*, 4, 187–207.
- 1965 "Die Entwicklung der Sprachlaute." *Phonetica*, 13, 95–98.
- 1972 *Structure of Bushman and its traces in Indo-European*. Wrocław: Zakład Norodowy im Ossolinskich.
- 1979 *Clicks, their form, function and their transformation or how our ancestors were gesticulating, clicking and crying*. Krakow, Nakl. Uniwersytetu Jagiellońskiego.
- 1983 "Supposed first words of apeman." In E. de Grolier, ed., *Glossogenetics: The origin and evolution of language: Proceedings of the International Transdisciplinary Symposium on Glossogenetics* (pp. 491–512). Paris: Harwood Academic Publishers.

Traill, A. and Vossen, R.

(1997, in press) "Sound change in the Khoisan languages: new data on click loss and click replacement." *Journal of African Languages and Linguistics*.

Turner, R. L.

1924 "The Sindhi recursives or voiced stops preceded by glottal closure." *Bulletin of the School of Oriental Studies*, 3, 301–315.

Varma, S.

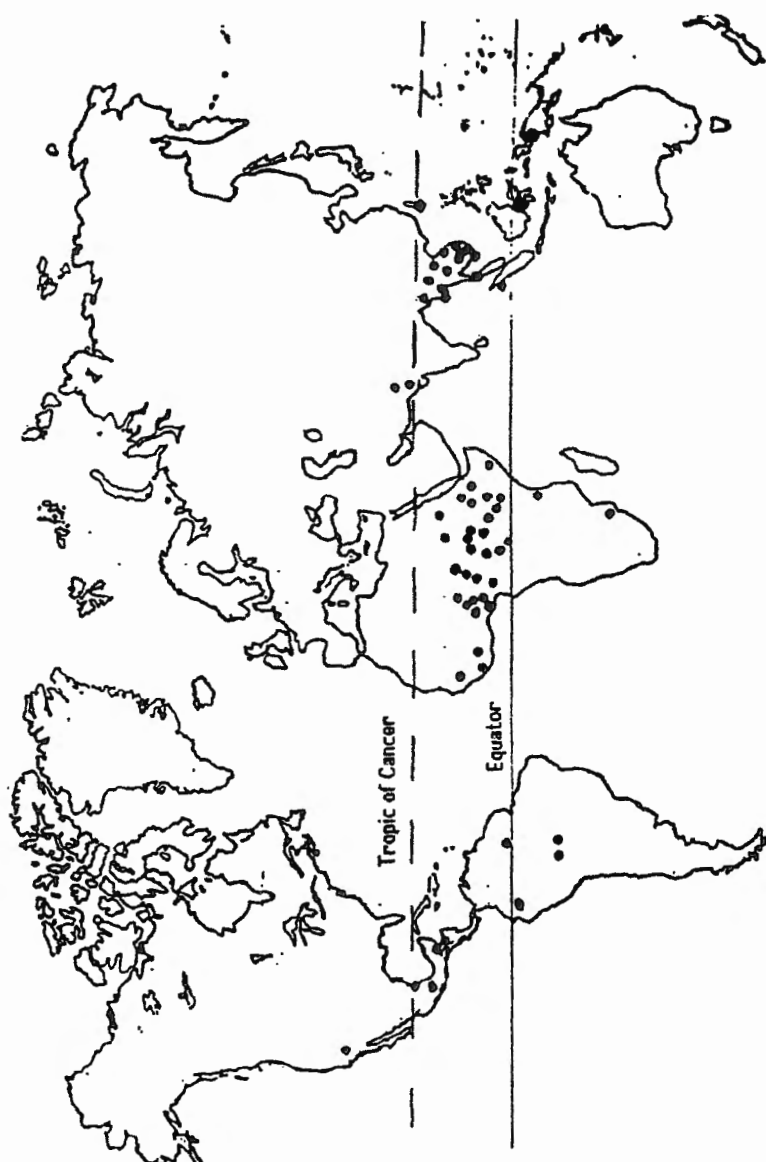
1948 *The Bhalesi dialect*. Calcutta, Royal Asiatic Society of Bengal.

Vendryes, J.

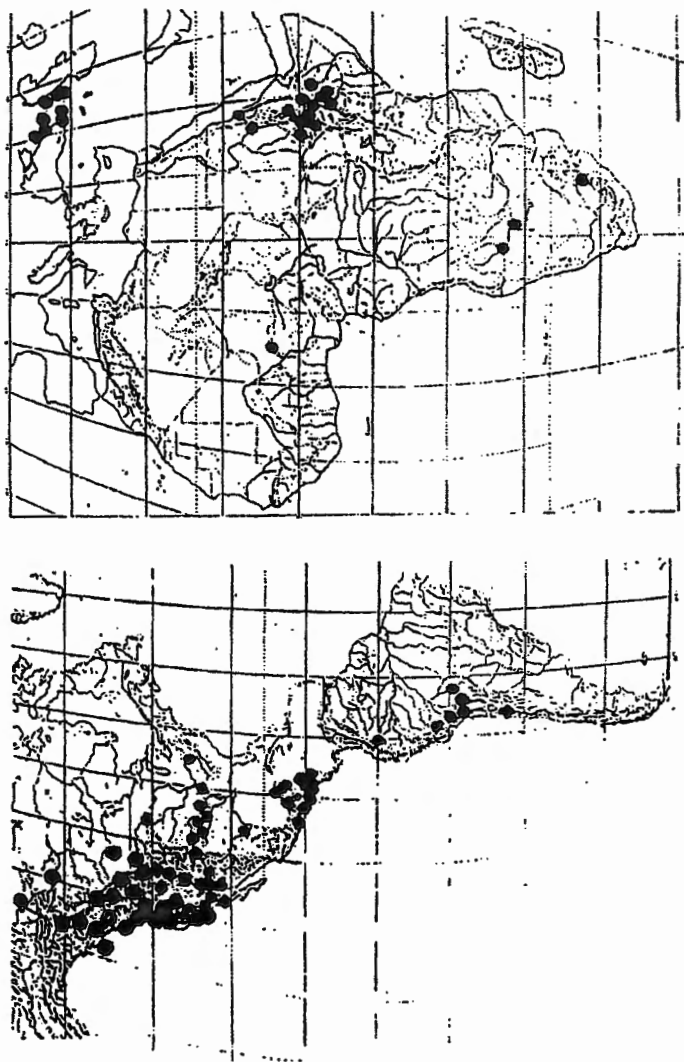
1923 *Le langage: Introduction linguistique a l'histoire*. Paris: La Renaissance du livre.

Vogt, H.

1936 *Esquisse d'une grammaire du Géorgien moderne*. Oslo: A., W. Brøggers Boktrykkeri.



Map 1: Languages with voiced implosives



Map 2. Languages with ejectives, in America, Africa and the Caucasus

## Pāṇini and the Distinctive Features

Madhav M. Deshpande<sup>1</sup>  
*The University of Michigan*

Several rules in Pāṇini's grammar presuppose the existence of a detailed system of phonetic features. Since some of these are metalinguistic rules -- they specify how regular rules in the grammar are to be interpreted -- this featural system can be said to form an essential component of Pāṇinian grammar. Realizing this, it comes as at least a mild shock to discover that Pāṇini has nowhere explicitly described this system.<sup>2</sup> Nor, apparently, did he directly bequeath knowledge of it to his earliest commentators, for they often show confusion as to its details. The whole dizzying edifice of Pāṇinian grammar seems thus to rest upon a hidden foundation!

Luckily, with a bit of careful digging, a good part of Pāṇini's groundwork can be laid bare. We can refer to the featural systems of the ancient phonetic treatises, the Śikṣās and the Prātiśākhya, and speculate that Pāṇini's system was probably similar, especially since the few featural terms he uses have parallels in these treatises. While one cannot be totally certain that these phonetic treatises, in their presently known form, are entirely pre-Pāṇinian, one can be certain that the tradition of phonetic analysis and description to which they belong is indeed pre-Pāṇinian. One can fruitfully compare these systems with Pāṇini's system, and one can also attempt to deduce Pāṇini's phonetic classifications, using the few clues that he has left for us.

Some of these lines of inquiry have been followed in the previous research on this subject.<sup>3</sup> In the present paper, I shall not go into the detailed, and sometimes convoluted, arguments made in

---

<sup>1</sup> An 'ancient' version of this paper was drafted some twenty years ago by my student James Bare, in consultation with me, when Bare was a graduate student at Michigan. However, after completing his doctoral work, he moved away from linguistics into other fields. I have rewritten this paper in view of later work, and hence I do not wish James Bare to be held accountable for the views expressed in this paper. I am indebted to him for his contribution to the earlier draft.

<sup>2</sup> Traditionally, the *Pāṇiniya-Śikṣā* is cited as the means by which Pāṇini filled this gap. It is now generally accepted, however, that this is in fact a later work which considerably post-dates Pāṇini, and is not known even to Patañjali.

<sup>3</sup> George Cardona (1969), Madhav Deshpande (1975), James Bare (1976), Paul Kiparsky (1991), and Robert Hueckstedt (1995), to mention some prominent studies.



the previous research publications by the various scholars, including myself, but will only refer to some of the conclusions of this previous research and move on to dealing with some theoretical considerations arising from the form of Pāṇini's featural system and the way in which Pāṇini used it in his grammar. Here, we shall only broadly touch upon the various issues, which have been discussed at length elsewhere, and perhaps need to be discussed further in greater detail.

The version of Pāṇini's featural system adopted by most of the later commentators can be shown, through various deductive arguments, to at least closely approximate the original system Pāṇini may have had. This is not surprising, considering the many centuries of critical examination it has been subjected to by the various commentators. However, the deductive arguments also seem to suggest that Pāṇini's original system was not totally identical with the one ascribed to him by the later commentators. This is indicated by the fact that the system ascribed to Pāṇini by these later commentators is at odds with certain considerations internal to Pāṇini's grammar. In order to present an approximate picture of Pāṇini's original system, it is best to begin with the system ascribed to him by the later commentators, and then to look at its various problem areas.

We shall take a close look at the phonetic system as presented in Bhaṭṭojī Dīkṣita's *Siddhānta-Kaumudī*<sup>4</sup> mostly on rules P.1.1.9 (SK no. 10), P.8.4.68 (SK no. 11), and P.8.2.1 (SK no. 12). Bhaṭṭojī recognizes three major featural parameters - points of articulation (*sthāna*), internal effort (*ābhyantara-prayatna*), and external effort (*bāhya-prayatna*). The term 'external effort' refers to a whole set of parameters rather than a single parameter.<sup>5</sup> The point-of-articulation features are assigned as follows: throat : *a, k, kh, g, gh, ṇ, h, ḥ*; root of the tongue: *\*k*; palate: *i, c, ch, j, jh, ñ, y, ś*; cerebrum: *ṛ, ṭ, ṭh, ḍ, ḍh, ṇ, r, ṣ*; teeth: *l, t, th, d, dh, n, l, s*; lips: *u, p, ph, b, bh, m, \*p*; nose: *m*.<sup>6</sup> The nasal stops also have the nose as a second point-of-articulation feature, however, the categorization of

<sup>4</sup>Here we shall not concern ourselves with the question of sources of Bhaṭṭojī's classifications, but shall deal with the classifications offered by Bhaṭṭojī as reflecting 'his' system in a synchronic sense.

<sup>5</sup>The distinction between 'internal' versus 'external' efforts is most probably not historically a Pāṇinian distinction, but is made explicitly by Patañjali in his discussion on P.1.1.9.

<sup>6</sup>The sound *\*k*, phonetically [x], and *\*p*, phonetically [ɸ], are contextual variants for *visarga, ḥ*.

nose as a point of articulation raises some important problems. The sounds *e* and *ai* are both said to have the throat and palate, and *o* and *au* both the throat and lips as points of articulation. The consonant *v* is said to have a double point of articulation - teeth and lips.

Internal effort is fourfold: contact, assigned to stops; slight contact, assigned to semivowels; openness, assigned to both spirants and vowels (except short *a*); and closeness, assigned to short *a*.<sup>7</sup> These classifications are displayed in the chart on the next page.

Bhaṭṭojī lists eleven external-effort features: openness (of glottis), closeness (of glottis), breath, tone, +voice, light breath (-aspiration), great breath (+aspiration), acute, grave, and circumflex. The unvoiced stops and the unvoiced spirants have the features of openness, breath, and -voice. The remainder of the consonants have the features of closeness, tone, and +voice. The unaspirated stops, both voiced and voiceless, the nasal stops, and the semivowels have the feature of light breath (-aspiration). The remainder of the consonants have the feature of great breath (+aspiration). Acute, grave, and circumflex are prosodic features of accent and apply to vowels only. Of the earlier features, the features of closeness of glottis and the resulting tone or resonance also apply to vowels. The vowels may also be said to have light breath. However, among vowels, all the instances share these features, and therefore, they do not distinguish one set of vowels from another set of vowels.<sup>8</sup>

Before venturing on a brief tour of the problem areas contained in this version, it would be well to restate exactly what it is that we mean by the term "problem area". Since we are only interested, for the moment, in discovering Pāṇini's original classificatory intentions, problem areas are those where the later classifications are suspicious, due to questions of consistency within the grammar. There are some areas where modern linguists would wish to question the rationale behind Pāṇini's featural classifications. However, for the present concern, these are not to be viewed as problem areas,

<sup>7</sup>The traditional grammarians, including Bhaṭṭojī, make the assumption that short *a* was a close sound in actual usage, but that within the confines of a grammatical derivation, it was considered to be - or perhaps pronounced - an open vowel, in order to achieve its necessary homogeneity (*sāvarṇya*) with *ā* by P.1.1.9 (*tulyāsyaprayatnam savarṇam*). The final rule of Pāṇini's grammar restores the close *a* in all derivations, before they become usable.

<sup>8</sup>A modern Indian Pandit, Jagadīśācārya Citrācārya, in his Sanskrit work *Sikṣāśāstram* (p. 13) says: *tatrodātānudātāsvaritās trayāḥ svarāṇām eva sarveṣāṃ / śeṣā aṣṭau vivārādāyo vyañjanānām eva*. This is clearly mistaken in denying vowels the features of glottal closure and resonance.

# Points of Articulation and Internal Efforts according to Bhaṭṭoji Dīkṣita

Points →	Throat	Root of Tongue	Palate	Cerebrum	Teeth	Lips	Nose
<b>Effort</b> ↓							
<b>Contact</b>	k kh g gh ṇ		c ch j jh ñ	ṭ ṭh ḍ ḍh ṇ	t th d dh n	p ph b bh m	ṇ ñ ṇ n m
<b>Slight Contact</b>			y	r	l, v	v	
<b>Openness</b>	h ḥ a e o ai au	×k	ś i e ai	ṣ ṛ	s ḷ	×p u o au	ṁ
<b>Closeness</b>	a						

if it is sufficiently clear that they reflect Pāṇini's classificatory intentions. A good example of this distinction is Pāṇini's supposed division of the internal-effort parameter so as to give spirants and vowels the same feature of internal effort, that of openness. Although this is puzzling to modern eyes, there is precedent for it in the phonetic treatises, and sufficient evidence within Pāṇini's grammar itself to suggest it as indeed being Pāṇini's intended classification.<sup>9</sup>

There is no lack of problematical classifications, however. The following list gives a brief treatment of these problematical classifications:

(1) Throat as the point-of-articulation feature for the *k*-class of stops. Most of the phonetic treatises have it as the root of the tongue, and there is some evidence within Pāṇini's grammar that supports this latter alternative.<sup>10</sup>

(2) Dual point-of-articulation features for nasal stops. This classification would probably give rise to problems within the grammar. An alternative, as found in some phonetic treatises, might have been to distinguish nasality along some other featural parameter, such as the articulator, rather than as a point of articulation. However, no clear evidence to decide between the different alternatives can be offered.<sup>11</sup>

(3) Dual point-of-articulation feature for *ṽ*. There is a good deal of evidence, though not uncontested, to classify *ṽ* not as labiodental, but as purely labial.<sup>12</sup>

(4) Dual point-of-articulation features for *e* and *o*. This classification leads to problems within the grammar. There is some evidence, though not totally conclusive, that Pāṇini intended these sounds to be treated as monophthongal, i.e., palatal only and labial only, respectively.<sup>13</sup>

(5) Same internal-effort feature for diphthongs and simple vowels. This also leads to problems within the grammar. A possible alternative, having precedent in the phonetic literature, is that Pāṇini set of *e*, *o*, *ai*, and *au* from the other vowels by means of

<sup>9</sup>Bare (1976), pp. 113-115.

<sup>10</sup>Bare (1976), pp. 126-130.

<sup>11</sup>Deshpande (1975), pp. 11-12; Bare (1976), pp. 130-140.

<sup>12</sup>For a detailed discussion, see Deshpande (1975-a) and (1981); Bare (1976), pp. 160-171; Cardona (1964).

<sup>13</sup>Deshpande (1975), pp. 138-140; Bare (1976), pp. 187-191; Cardona (1983).

another internal-effort feature. Again we have no conclusive evidence within the grammar.<sup>14</sup>

The system which emerges, however obscurely, from these considerations obviously bears strong resemblance to the point-of-articulation/mode-of-articulation phonetic charts of the pre-generative era. Non-binary features, although lacking a certain cybernetic efficiency, do not seem to have interfered with the phonological roles Pāṇini chose for his features to play, though they may have limited the extent of these roles, as we shall see below.

The system as such cannot be called innovative, since its general shape as well as almost all its particular classifications have close parallels in the phonetic treatises.<sup>15</sup> Indeed, at times one feels that Pāṇini is going through a great deal of effort to preserve some inherited classification, even while realizing the problems it posed to the phonological operation of the grammar. A good example of this is the lack of distinction between vowels and spirants along the parameter of internal effort. It is almost as if Pāṇini, like modern phonologists, wanted to make sure that his featural system was basically phonetically motivated; i.e., decided upon a purely phonetic examination of sounds before reference was made to its convenience in describing the phonological connection.

If it is true that Pāṇini did choose, for whatever reasons, a featural system that was quite closely bound by the precepts of traditional Indian phonetics, two questions come to mind: (1) How heavy could the functional load of such a system be within a grammar that was truly innovative? (2) What limitations, both theoretical and mechanical, did the utilization of such a system impose upon the grammar?

The answer to the first question to a certain extent vindicates traditional Indian phonetics as well as Pāṇini's apparent conservatism, for the scope of application permitted by his choice of this particular system is actually quite broad. Pāṇini's use of features is basically twofold: (1) He defines certain types of natural classes in terms of features so that these classes may be utilized in rule statements. (2) He uses features as a means by which many single-sound substitution rules can be combined into a unified, more general class-substitution rule.

Pāṇini uses two kinds of featurally-defined natural classes. In the first of these, he incorporates a term derived from a point-of-

<sup>14</sup>Bare (1976), pp. 191-192; Cardona (1983).

<sup>15</sup>For a general description, see Allen (1953) and Bare (1976).

articulation feature in the rule itself in order for it to represent all sounds having that particular feature. This usage occurs in only three rules, however, and thus has somewhat limited significance. The terms used are 'labial' (*oṣṭhya*), 'dental' (*dantya*), and 'cerebral' (*mūrdhanya*).<sup>16</sup>

The second kind of natural class results from Pāṇini's definition of the term 'homogeneous' (*savarṇa*), through which sounds having a common point-of-articulation feature and a common internal-effort feature are grouped together as homogeneous sounds. By this means short and long vowels are grouped together as a class (e.g., *a* and *ā*, *i* and *ī*, etc.); likewise stops sharing a particular point-of-articulation feature (e.g., *k*, *kh*, *g*, *gh*, and *ṅ*). Classes of this type are usually represented in rules not directly by featural terms, but by specified tokens to represent the types or classes. For example, *a* stands not only for itself, but for the whole class of eighteen homogeneous sounds:

*a ā ā̃ á â ã à â ã̃*  
*ã ã̃ ã̃̃ ä ä̃ ä̃̃ ã̃ ã̃̃ ã̃̃̃*

Similarly, the consonant *k*, with a marker *U* attached to it, i.e. *kU*, represents the group of homogeneous consonants, i.e. *k*, *kh*, *g*, *gh*, and *ṅ*.<sup>17</sup>

This is the extent of Pāṇini's employment of featurally defined natural classes. He of course needed to have a means by which other classes than these could be represented in rules, and for this he devised an ingenious indexing device which is not directly based on features. This device lists the sounds of Sanskrit in a certain order with various index sounds inserted at key points, thus:

1. *a i u Ṇ*
2. *r ḷ K*
3. *e o ṅ*
4. *ai au C*
5. *h y v r ṭ*
6. *l Ṇ*
7. *ñ m ṇ n ṅ M*
8. *jha bha Ṇ*

<sup>16</sup>The rules are respectively P.7.1.102, P.7.3.73, and P.8.3.55.

<sup>17</sup>This procedure of representation of the homogeneous sounds by a specified token is prescribed by P.1.1.69. For a detailed study of the notions of *savarṇa* and *savarṇa-grahaṇa*, see: Deshpande (1975).

9. *gh dh dh Ṣ*
10. *j b g ḍ d Ṣ*
11. *kh ph ch ṭh th c ṭ t V*
12. *k p Y*
13. *ś ṣ s R*
14. *h L*

These fourteen sound-strings are the famous *Śivasūtras*, traditionally believed to have been given to Pāṇini by Śiva.<sup>18</sup> When Pāṇini wanted to refer to a certain class by means of this device, he used what can be called a shortform (*pratyāhāra*), consisting of any one of the listed sounds (except the marker sound given here as upper-case letters) as the first sound, together with any one of the subsequent marker sounds. Such a shortform represents all the sounds beginning with the first sound of the shortform upto the given marker sound, excluding the marker sounds themselves. For example, *aC* is a shortform used to represent *a, i, u, ṛ, ḷ, e, o, ai*, and *au*. The shortform *ñ(a)M* represents all the nasal stops.

It can be ascertained by an examination of the grammar that most of the classes represented by means of this device are what we would want to call natural classes, and could therefore be represented by means of some feature-based device. The question then, of course, is why Pāṇini chose to represent some classes by means of features and not others. One might mistakenly expect the answer to have something to do with the limitations of Pāṇini's featural system, i.e. that it was not versatile enough to allow representation of other types of classes. This, however, is clearly not the case. The system, in as much as we can pin it down, would have allowed featural definition of nearly every natural class actually represented by means of the indexing lists or shortforms.

The real answer evidently is that since Pāṇini's grammar, as all other scholarly works of the time, was designed for oral transmission, brevity of rule formulation was a priority consideration. That is, even though *jhaṢ* was perhaps not as theoretically elegant a class marker as 'voiced aspirate stop' or its Sanskrit equivalent, it certainly was shorter and easier to memorize as part of a rule. In addition, although featural reference was possible by means of the featural system Pāṇini chose, in some cases such reference would

<sup>18</sup>For a detailed study, see George Cardona (1969). The most recent study of these strings is Kiparsky (1991). Other studies of these strings are listed by Cardona and Kiparsky.

have been especially unwieldy, due, among other things, to the basic nonbinary set-up of the system. For example, the class referred to by Pāṇini as *śaR* - *ś*, *ṣ*, and *s* -- could indeed have been defined featurally by some sort of listing of each segment's features, e.g., palatal, cerebral, and dental open unvoiced aspirates, or some such. This, in fact, is one of the main drawbacks of nonbinary systems - every feature along a given parameter is assumed to have equal affinity with all the other features along that parameter, so that classes formed from a subset of all the features along the parameter are marked, through the unwieldy necessity of listing the individual segments' features, as being nearly as unnatural as a class of unrelated segments. That is, the class *ś*, *ṣ*, and *s* is nearly as difficult to describe featurally as the class *a*, *y*, and *d*. One may feel that one could featurally describe the class of *ś*, *ṣ*, and *s* by the description *aghoṣa ūṣman* 'voiceless spirants'. However, this will also include the sounds *h*, *\*k*, and *\*p*. To exclude these, one indeed needs to make some unwieldy efforts. It is thus probable that the shape of the system chosen by Pāṇini did put some constraints on his use of features in defining natural classes, but considerations of brevity were still probably the more important ones for him. That is, even had the system chosen been binary, the indexing list would most likely still have been the prevalent mode of class representation.

For those classes where Pāṇini did have recourse to features, the rationale seems to have been either impossibility of reference by means of the indexing list (as for classes based on point-of-articulation features) or realization of the fact that certain classes appeared in rules much more often than any individual members or subsets of those classes (e.g., a rule is much more likely to deal with all the varieties of *i*--*i*, *ī*, *ī3*, etc. -- than any single variety or subset of varieties).<sup>19</sup>

Thus, even though featurally-based classes take something of a secondary role in Pāṇini's grammar, the important thing is that he realized that features could be useful in this function. His use of shortforms in preference to features is indicative not of theoretical naiveté, but rather of practical craftsmanship and regard for the milieu in which he found himself working.

<sup>19</sup>Since *i* in a rule represents the whole homogeneous class, when Pāṇini did want to refer to just *i* or just the short varieties, he had to have recourse to a special device in order to do so. Through the use of the marker *T*, a sound was enabled to represent homogeneous sounds of the same length. For a detailed study of this procedure, see Deshpande (1972).



Apart from class-formation, Pāṇini uses features as a means by which many single-sound substitution rules can be combined into a single class-substitution rule. He achieves this through a rule-interpretation device known in Sanskrit as *āntaratamya*, which may be literally translated as 'substitution of the most similar'. This device provides a criterion for deciding which of a class of possible substitutes should be chosen for a given substituendum -- the criterion being greatest featural similarity or proximity.<sup>20</sup> For example, if a rule states that the class *i, ī, u, ū, r, ṛ*, and *l* is to be replaced by *y, v, r*, and *l*, when followed by a vowel, substitution of the most similar would operate to specify *y* as the proper substitute when *i* or *ī* is the substituendum, *v* when *u* or *ū* is the substituendum, etc. The "greatest similarity or proximity" in this case would of course result from the fact that *y* is featurally more similar to *i* or *ī* than is any of the other possible substitutes, the point of articulation feature 'palate' being held in common. Thus, instead of having to state four separate rules --

$$i, \bar{i} \Rightarrow y / -V$$

$$u, \bar{u} \Rightarrow v / -V$$

$$r, \bar{r} \Rightarrow r / -V$$

$$l \Rightarrow l / -V$$

Pāṇini, by means of this device of maximal featural proximity, not only could satisfy his self-imposed conditions of economy of rule statement, but was also able to capture a linguistically significant generalization that the other means of statement would have missed.<sup>21</sup>

It is interesting to note that modern phonology uses features in rule statements to fill these same two basic needs -- the representation of natural classes and the combination of analogous processes into a unified, general rule. The difference is that Pāṇini seems not to have seen these two as aspects of the same thing, as does modern

<sup>20</sup>Bare (1976), pp. 99-111, discusses the various problems involved in our understanding of this procedure.

<sup>21</sup>This is a very diluted presentation of the operation of P.6.1.77 (*iko yan aci*). The rule has provoked many controversies and we cannot go into those details here. For details see: Deshpande (1981), p. 61, and Cardona (1980-81), pp. 396ff. Robert Hueckstedt's 1995 monograph *Nearness and Respective Correlation* is devoted entirely to the considerations arising out of the interpretation of this rule.

phonology, where a single device -- rule statement *in terms of features* -- meets both these needs. This difference is theoretically crucial. Pāṇini saw features as facilitators for rule statements, but felt the statements themselves to properly concern segments, whereas modern phonology sees rule statements as being about features rather than segments. The devices used by Pāṇini show that he came extremely close to making this connection himself; the fact remains, however, that he never took the final step. This may indeed have had something to do with the prevalent oral tradition of the time, where considerations of brevity and mnemonic ease were probably just as important as the theoretical ones. Certainly most of Pāṇini's phonological rules would, from this point of view, have been more unwieldy if stated in pure featural terms. It is also the case that Pāṇini could express everything that he needed to express within his grammatical system by means of the devices he adopted. He did not concern himself with extra-grammatical topics like historical change, where rule statement in terms of features might have enabled greater generality of expression, and also may have provided a greater insight into the causation of sound-change. If this is valid, we can say that Pāṇini, through the devices in which he utilized features, got as close to the notion of rule statement in terms of features as his times and focus allowed him to.

Even though the extent to which Pāṇini utilizes features in his grammar is somewhat more limited than in modern phonological practice, the mere fact of utilization led him to struggle with a problem that has plagued phonologists to this day - how to reconcile phonetics and phonology. That is, how does one relate, through a single featural system, the phonetic "facts" and the phonological requirements of a grammar when the two often seemingly demand different featural classifications for the same sound.

The phonological demands upon Pāṇini's featural system were directly determined by the ways in which Pāṇini chose to utilize the system in his grammar. These demands, articulated as ideal expectations, are basically of two types: (1) Where Pāṇini uses features to define classes, the implicit demand upon the system is that all the resultant classes be phonologically desirable, i.e., useful in the statement of rules, and, conversely, that all the relevant desirable classes result from the basic featural definition. (2) Where Pāṇini uses featural proximity as a means for determining which of a class of substitutes is to be chosen for a given substituendum, the implicit demand upon the system is that it provide a basis whereby the single

desired substitute can be identified as the "most similar or proximate" to the given substituendum.

Pāṇini seems to have had more trouble dealing with the demands imposed by (1) than with those imposed by (2). The most glaring problem was probably that vowels and spirants sharing the same point of articulation were classed together as homogeneous sounds by virtue of their sharing the same feature of internal effort. Since pairs like *h* and *a* (and *ḥ* and *ā*), *ś* and *i*, etc., were obviously undesirable from a phonological point of view as pairs of homogeneous sounds, Pāṇini was constrained to attach as a condition to his basic featural definition of homogeneity a statement to the effect that homogeneity could not obtain between vowels and consonants.<sup>22</sup>

We can identify at least three other areas where the phonetic conditions for homogeneity seem to be met and yet homogeneity must be considered undesirable from a phonological point of view. Unfortunately, in none of these areas can we be certain how Pāṇini himself chose to resolve the dilemma, due to lack of evidence within the grammar. The first of these areas is that of the diphthongs. We do not know Pāṇini's exact phonetic intentions in this area, but we do know that the classifications found in some of the phonetic treatises would have led to solutions of these problems. We also know that Pāṇini intended no homogeneity among the sounds *i*, *e*, and *ai*, or *u*, *o*, and *au*. There is some evidence, as stated above, that Pāṇini used phonetic criteria to distinguish *ai* and *au* from the others, but we really have no way of knowing how he dealt with *i* and *e*, or with *u* and *o*.<sup>23</sup>

A possible area of this type is that having to do with *h* and *ḥ*. We know that if Pāṇini followed the classifications of the great majority of phonetic treatises here, these sounds would meet the phonetic criteria for mutual homogeneity. It is also the case that such homogeneity must be judged undesirable from a phonological standpoint. Here again we do not know how Pāṇini chose to handle this. A check through the grammar shows that even if Pāṇini allowed the theoretical problem of undesirable mutual homogeneity to remain, no serious practical malfunction of the grammar would

<sup>22</sup>P.1.1.10 (*nājjhalau*). The exact interpretation of this rule has led to a great deal of controversy from the earliest period of the Pāṇinian tradition. For details, see Madhav Deshpande (1975), pp. 61-69. Also see: Bare (1976), pp. 95, 97, 113-15.

<sup>23</sup>For further discussion, see Madhav Deshpande (1975), pp. 138-140, Bare (1976), pp. 185-192, and Cardona (1983), pp. 13-31.

result, due to the almost mutually exclusive contexts, both phonetic and phonological, in which *h* and *ḥ* can appear.<sup>24</sup>

Pāṇini probably faced another dilemma when dealing with the nasalized sounds -- the nasal stops, semivowels, and vowels. Certain early phonetic schools conceived of these sounds as having two points of articulation, an oral one and the nose.<sup>25</sup> However, such a phonetic classification would have led to two types of phonological complication with respect to homogeneity. On the one hand, Pāṇini demonstrably wanted mutual homogeneity between nasal and nonnasal stops having the same oral point of articulation (likewise with semi-vowels and vowels); on the other hand, there is no indication that Pāṇini wanted mutual homogeneity among all the nasal stops (or semivowels or vowels). Again we cannot be sure how Pāṇini chose to handle this. He might have implemented a non-problematical phonetic classification by giving these sounds only oral point-of-articulation features and then accounting for the nasality distinction along a featural parameter that did not figure into the definition of homogeneity. On the other hand, he may indeed have adopted the problematical phonetic classification and chosen to deal with the problem by putting an implicit constraint on the phonological operation of the grammar, possibly to the effect that the nose as point of articulation feature was to be excluded from calculations of homogeneity.<sup>26</sup>

There is at least one other area where the phonological requirements of homogeneity and the phonetic "facts" are apparently in conflict -- the classification of short *a*. Here Pāṇini realized that homogeneity between short *a* and long *ā* was desirable, since on a phonological level they behaved as similarly as, say, short and long *i*. However, short *a* was phonetically classified with a different internal effort feature than that of the remainder of the vowels, including long *ā*, and consequently did not meet the phonetic requirements for homogeneity with long *ā*.<sup>27</sup> Pāṇini's solution to this is interesting and unusual. He classified short *a* with the same feature of internal effort as that of long *ā* so that the desired homogeneity would be obtained, and then he formulated a rule, which in effect applied at the end of every derivation, to return the classifica-

<sup>24</sup>Bare (1976), pp. 144-153.

<sup>25</sup>Bare (1976), p. 119.

<sup>26</sup>Deshpande (1975), pp. 11-12; Bare (1976), pp. 130-140.

<sup>27</sup>The short *a* was evidently phonetically [ə] in Pāṇini's time, and therefore qualitatively as well as quantitatively different from *ā*, phonetically [aa]. For a detailed discussion of the phonetics of short *a*, see Deshpande (1975-b).

tion of short *a* to its more proper phonetic value. In instituting a different classification at each level, Pāṇini hit upon a way or expressing rather than avoiding the problem, and thus found a solution that was both descriptively and theoretically pleasing.

The phonological requirements of *āntaratamya* ('substitution of the most proximate') would seem to be such that a number of problem areas might be expected here also, since phonetic and phonological similarity or proximity are not always equivalent. Many of these potential problem areas never seem to manifest themselves, however, probably due to controlling this device through individual rule statement. Despite this, there are a few items of interest that deserve mention.

It is indeed a possibility that Pāṇini was led to his labial point-of-articulation for *v* through phonological considerations having to do with *āntaratamya*, i.e., so that *v* would be more similar to *u* than to any other vowel.<sup>28</sup> Similar considerations might have contributed to the cerebral classifications for *r* and *ṛ*, the usual classification for these sounds in the *Prātiśākhya*s being non-cerebral.<sup>29</sup> It seems doubtful, however, that Pāṇini, usually rather conservative in his other phonetic classifications, would have been willing to make such shifts in these classifications unless there were a factual basis for them.

The picture emerges, from all this, of a pioneer phonologist cautiously feeling his way through this territory. From our modern perspective, we can perhaps detect certain limitations in Pāṇini's approach. But, surely, even the fact that Pāṇini's featural system, both in shape and utilization, can be discussed in terms that are not too alien to modern sensitivities is little short of remarkable. How much more so when we come across solutions in Pāṇini that can still stand up to the latest theories in phonology.

## References

Allen, W.S.  
1953

*Phonetics in Ancient India*. London Oriental Series  
1. London: Oxford University Press.

<sup>28</sup>Deshpande (1975-a), Bare (1976), pp. 160-171.

<sup>29</sup>Bare (1976), pp. 171-2; Deshpande (1978), pp. 282-285.

Bare, James

1976

*Phonetics and Phonology in Pāṇini: The System of Features Implicit in the Aṣṭādhyāyī*, Natural Language Studies, No. 21. Ann Arbor: The University of Michigan Phonetics Laboratory.

Cardona, George

1964.

"The Formulation of Pāṇini 7.3.73". In *the Journal of the Oriental Institute*, Baroda, Vol. 14, pp. 38-41.

1969

*Studies in Indian Grammarians I, the Method of Description Reflected in the Śivasūtras*. Transactions of the American Philosophical Society, New Series, Vol. 59, Pt. 1. Philadelphia: American Philosophical Society.

1980-81

"On the Domain of Pāṇini's Metarule 1.3.10". In *Adyar Library Bulletin*, Vols. 44-45, pp. 394-414.

1983

"Phonetics and Phonological Rules in Grammars." First lecture in *Linguistic Analysis and Some Indian Traditions*. Shripad Shastri Deodhar Memorial Lectures, Series 1. Pune: Bhandarkar Oriental Research Institute.

Deshpande, Madhav M.

1972

"Pāṇinian Procedure of Taparakaraṇa: A Historical Investigation." In *Zeitschrift für vergleichende Sprachforschung*, Band 86, Heft 2, Göttingen, pp. 207-254.

1975

*Critical Studies in Indian Grammarians I: The Theory of Homogeneity [Sāvarṇya]*. The Michigan Series in South and Southeast Asian Languages and Linguistics, No. 2. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.

1975-a

"Phonetics of V in Pāṇini." In *Annals of the Bhandarkar Oriental Research Institute*, Vol. 56, pp. 45-65.

1975-b "Phonetics of Short A in Sanskrit." In *Indo-Iranian Journal*, Vol. 17, pp. 195-209.

1978 "Genesis of Rgvedic Retroflexion: A Historical and Sociolinguistic Investigation." In *Aryan and Non-Aryan in India*, edited by Madhav M. Deshpande and Peter E. Hook. Michigan Papers on South and Southeast Asia, No. 14, pp. 235-315. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.

Hueckstedt, Robert A.

1995 *Nearness and Respective Correlation : A History of the Interpretations of Aṣṭādhyāyī 6.1.77 : iko yaṇ aci*. Wiesbaden: Otto Harrassowitz.

Jagadīśācārya Cītrācārya

1969 *Śikṣāśāstram* (in Sanskrit). Publication 1. Bahraich, U.P., India: Bālārka Veda Mandira.

Kiparsky, Paul

1991 "Economy and the Construction of the Śivasūtras." In *Pāṇinian Studies, Professor S.D. Joshi Felicitation Volume*, edited by Madhav M. Deshpande and Saroja Bhate, Michigan Papers on South and Southeast Asia, No. 37, pp. 239-261. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.

*Mahābhāṣya* by Patañjali

Edited by F. Kielhorn. Rev. ed. by K.V. Abhyankar. Vol. 1, 1962; Vol. 2, 1965; Vol. 3, 1972. Pune: Bhandarkar Oriental Research Institute.

*Siddhānta-Kaumudī* by Bhaṭṭojī Dīkṣita

with the commentary *Tattvabodhinī* by Jñānendra Sarasvatī. Edited by Vasudeva Shastri Panashikar. 7th Edition. Bombay: Nimaya Sagara Press, 1933.

## Does Altaic Exist?

Joseph H. Greenberg  
*Stanford University*

Since the writings of Clauson, and more recently Doerfer, it appears that most specialists in the Altaic languages no longer believe that the three groups of traditional Altaic, namely Turkic, Mongolian, and Tungusic, are related; their resemblances are to be attributed to borrowing, or in some cases to accident or sound symbolism.

The term 'traditional Altaic' is here used purposely, that is, without reference to Korean, Japanese, or for that matter Uralic<sup>1</sup>. This is not because I believe that the Altaic languages are genetically isolated. In fact, in my view (Greenberg 1987a: 332), they belong to a much larger grouping, Eurasiatic, along with other languages besides those just mentioned above. Moreover, considerations deriving from these wider connections will figure in some instances in an essential way in the following discussion.

There are two separate questions involved here. Are the Altaic languages related to each other? If they are, do they constitute a valid genetic grouping, that is, a set of languages which have a single exclusive common ancestor, Proto-Altaic, which gave rise to three groups of languages and no others?

I believe that the answer to the first question, that of mere relationship, is overwhelmingly positive. That to the second is more difficult, but on the balance I rather strongly endorse a positive answer here also.

Recently in several publications, Miller (1991a, 1991b) has defended the traditional view. His arguments are largely phonological, especially the existence of two reconstructed pairs of liquid phonemes  $l_1$ ,  $l_2$ ,  $r_1$ , and  $r_2$ , which within Altaic are only distinguished in non-Chuvash Turkic. Miller believes that  $l_1$  and  $l_2$  have separate reflexes in Japanese. There are also instances in which Turkic merges a number of phonemes in  $j$ , namely  $d$ ,  $j$ ,  $n$ , and  $n^y$ . In such instances in order to account for the usual anti-Altaicist scenario in terms of borrowing from Turkic into Mongolian (with some reverse borrowing) and then from Mongolian into Tungusic, the borrowing has to be pushed back to a time so early that it becomes indistinguishable from Proto-Altaic, that is, when Turkish still distinguished  $d$ ,  $j$ ,  $n$ , and  $n^y$ , and all the Altaic languages outside of non-Chuvash Turkic displayed a difference between  $l_1$  and  $l_2$  as well as  $r_1$  and  $r_2$ . At such a time the languages would all have had a sound system which is identical with that reconstructed by Ramstedt, Poppe, and others for Proto-Altaic.

---

<sup>1</sup> It seems clear to me that languages like Korean, Japanese, and Uralic stand apart from traditional Altaic. Thus, Poppe (1960: 8), who includes Korean, shows it as a separate branch from the rest of Altaic, and it figures comparatively infrequently in his etymologies.



Miller also alludes to the cogency of the grammatical data regarding verb derivation in Ramstedt (1912) and Poppe (1973). I agree with him on all of this, but I believe that he has omitted the most powerful evidence of all, that based on personal, demonstrative, and interrogative pronouns.

This material is, of course, familiar, but the anti-Altaicists have, as will be shown, carefully avoided presenting it in a coherent way, and where they have, have sought to explain it away in an unconvincing fashion as the result of factors other than common genetic inheritance.

I will begin with the first and second person pronouns. In the first person singular in non-Chuvash Turkic, some languages, e.g. Osmanli Turkish, have nominative singular *ben* and a stem *ben-* which, except for an internal variation in the dative (*banā*), is found in all the oblique cases. Most Turkic languages, however, have *men* rather than *ben*, and all have *-m* as the first person singular marker in verb forms. The fundamental form then is *me-n*, in which *-n* (often called pronominal *n* by Altaicists) has as its original function a mark of the oblique, ultimately of genitive origin. In non-Chuvash Turkic, this *-n* has spread analogically to the nominative. In Chuvash, however, which represents a separate branch of Turkic, this did not occur. The nominative here is *e-pě* in which *e* is a deictic element, and the oblique stem is *man-*.

This irregular alternation between nominative and oblique recurs in Mongolian in which the nominative is *bi* and the genitive *min-u* and Tungusic, e.g. Evenki, with nominative *bi* and genitive *min-i*. The forms *men* and *min* are much more widespread than Altaic, including Uralic (e.g. Finnish *minä* 'I') and Indo-European. Indo-European appears here as an important link in this chain. On the basis of Baltic, Slavic, and Indo-Iranian, Szemerényi (1970: 197) reconstructs *\*mene* for the genitive. In Baltic and Slavic, the form in *-n* has been extended to all the oblique cases as in Altaic.

The Indo-European evidence is important because it provides a confirming instance for the oblique case function of the form in *-n*. This is presumably the same *-n* which occurs in the oblique cases of *r/n* stems<sup>2</sup>. The Indo-European independent nominative is a suppletive form but different from that of Altaic, namely *e-g(h)o-m*, whose most closely related form in Eurasiatic is Chukchee *i-g m/e-g m* (vowel-harmony variants) 'I' (cf. *i-g t/e-g t* 'thou'). Forms without the initial vowel occur as bound objects).

Returning to Altaic, it is clear that the probability of an irregular alternation such as *bi/men* occurring three times by accident is infinitesimal. That it should be borrowed twice is also utterly improbable. One has literally to scour the earth to find a few instances of a borrowed pronoun, much less an entire irregular

<sup>2</sup> The oblique *-n*, and indeed all the grammatical elements here were discovered by the Nostraticists. See especially the tables in Illič-Svityč (1971: 6-18). I discovered these independently at a time when I was not aware of Nostratic. In some instances, of course, I have found additional support, especially in languages not included in "classical Nostratic," but often accepted now as Nostratic, e.g. Chukchi-Kamchatkan and Eskimo-Aleut.

alternation in pronouns. By itself it is enough to show that the Altaic languages are related, moreover the specific innovation of *bi* in the nominative is confined to these languages. Therefore it can be considered a shared common innovation within Eurasiatic that contributes to the establishment of traditional Altaic as a valid genetic entity.

How is this evidence treated by Clauson and Doerfer, the two leading exponents of the anti-Altaicist position? It is ignored where possible. In Clauson (1969: 38), which applies glottochronology to the Altaic problem, discussion is unavoidable since 'I' is part of the glottochronological list. He seeks to argue away the three-fold resemblances, indicated by italicized entries, among Old Turkish, Old Mongolian, and Manchu, the three languages he utilizes in his study as follows:

It is known (but has not been explained up to now) that there are phonetic resemblances between personal pronouns in languages which are completely unconnected with each other, e.g. between 'mine', German *mein* and the Turkish genitive *menin* (from *ben*) and Mongolian *minö* [sic!] from *bi*; between Latin *tu* and Mongolian *či* (\**ti*). The phonetic resemblances between Turkish, Mongolian, and Tungus-Manchurian in regard to these lexical items cannot be therefore recognized as probative.

This reasoning, which is very common, is to deny the significance of a resemblance because it is found somewhere else. This was used by Michelson against Sapir in regard to *n* first person, *m* second person in Algic because it occurs in so many other Amerind languages. It would be just as logical to deny the significance of the resemblance between English 'mine' and German *mein* because it also occurs in Mongolian. One has to pursue the full distribution of these forms. As soon as one gets to Sino-Tibetan or Nilo-Saharan, or many others, it ceases. Both the Nostraticists and I include Indo-European and Altaic in the same group. [For another perspective on widespread similarities in pronominal systems, see Rhodes, this volume. -Eds.]

In addition, Clauson, by simply using the nominative as the translation form for the glottochronological list, fails to consider the agreement between Mongolian and Tungusic in the *bi/min-* alternation, and by not including Chuvash does not have to account for the threefold agreement in an irregularity among the three branches of Altaic.

And what of the second person singular pronouns? They are not discussed at all. Clauson unaccountably does not italicize Old Turkish *sen* and Manchu *si* as resemblances to be explained, or rather explained away, in spite of their complete parallelism with Old Turkish *ben* and Manchu *bi*. Old Mongolian *tere*, Manchu *tere* 'this' are italicized but passed over without comment.

Doerfer in general fails to discuss grammatical resemblances, but in his *Mongolo-Tungusica* (1985: 2), he says the following about the first person

singular pronoun:

Indeed, even such an apparently clear comparison as Mongolian *bi* - Tungus *bi* is not convincing on closer examination, since the Mongolian forms (on account of the plural *bi-da*, cf. *e-de* 'these', *te-de* 'those') goes back to *bɪ*. A typical case of sound symbolism (*Elementarverwandschaft*), surface resemblance, but without the possibility of a connection by sound correspondence.

What Doerfer is saying is that Mongolian *i*, which has two sources in a system of back-front vowel harmony, must derive from a high *back* vowel, not a high front vowel, because of the vowel of the second syllable *-da* which is a back vowel.

What Doerfer fails to point out is that Mongolian *bida* is a first person inclusive plural. Now it is a worldwide typological fact that where there is a first person inclusive/exclusive distinction in the plural, the exclusive, when analyzable, is the plural of the first person. This is so in Mongolian, in which the first person is *ba*, with a perfect parallelism between the first and second persons, *bi:ba* = *či* < \**ti:ta*.

On the other hand the first person inclusive is either a separate form unlike either the first or second person singular, or it is a combination of the two like Tok Pisin *yu-mi*. Hence *bi-da* is very likely a compound of singular *bi* with *ta* second plural. In compounds vowel harmony need not apply. A parallel situation is found in Tungusic, in which most languages have a first person plural inclusive/exclusive distinction in which the exclusive is the plural of the singular. The same parallelism reigns here as in Mongolian between the first person and the second person, e.g. Evenki *bi:bu* = *si:su*. The first inclusive is here even more obviously a compound, e.g. Evenki *mi-ti*, *mi-t* (Tsintsius 1949: 270-1).

Note also that Doerfer fails to mention the striking parallelism between the nominative and oblique stems in the first person among Mongolian, Tungusic, and Chuvash. We are to believe that Mongolian *bi* here is not cognate with the Tungusic and Turkic forms in spite of the agreement between them in parallel irregularities. Characteristic also is Doerfer's resort to sound symbolism. This is done without any supporting evidence. Surely *b-* is not particularly frequent as a first person singular in languages of the world, nor is there any plausible support in sound imitation or other sources of *Elementarverwandschaft*.

Finally, it should be noted that violations of back-front vowel harmony are not uncommon in Uralic, a universally accepted family, and in etymologies which are obviously valid on other grounds. As late as 1910, Szinnyi, in his reconstruction of Proto-Finno-Ugric, resorted to a kind of majority rule to determine whether back or front vocalism was the original type in Proto-Finno-Ugric. Even now there are uncertain instances. A parallel situation exists in Turkic. As noted by Radloff (1882: 84) there are variations in stem vowels

without any demonstrable cause. In fact there is an article by Dmitrijev on this topic, (1955: 115) in which he observes that sporadic alternations in the same root of vowels of the front and back series is frequent in individual Turkic languages.

Another one of the very few grammatical etymologies in Doerfer (1985: 27) is his no. 66, the interrogative stem *ya-* of Mongolian and Tungus. He admits that it "behaves like a genetically related word." Once more he resorts to "sound symbolism" and again his only support is Indo-European *\*jo*. But this is a widespread Eurasiatic interrogative (cf. Greenberg 1987b). Once more we have the *ad hoc* resort to a highly implausible sound symbolic argument without any serious documentation.

Finally, what of the second person pronouns? They are passed over in complete silence. Doerfer, like Clauson, believes that Mongolian borrowed massively from Turkic, and then Tungusic from Mongol. He is clearly disturbed by the existence of certain etymologies common to Turkic and Tungusic and devotes a section to them (1987: 238-41), but he fails to mention the most glaring instance of all, the agreement of Turkic and Tungusic in an *s* second person as against Mongol *t*. Of course, if I am right in my discussion of the Mongol and Tungusic first person inclusive pronoun, *t* would also occur in Tungusic, but in a quite different context. Both *s* and *t* are widespread second person Eurasiatic pronouns. For example, we find Indo-European *t* in the independent pronoun and plural verb endings and *s* as a singular verb suffix.

In general there are a considerable number of other grammatical markers common to all the Altaic branches, most of them entirely ignored by Doerfer. However, virtually all these are found in other branches of Eurasiatic. The number of these as well as the lexical evidence makes the relationship of the Altaic languages a certainty. However, the distinctness of Altaic as a valid subgroup, which is most conspicuously supported by the *bi/min* alternation in the first singular pronoun requires further assessment, a task not undertaken here.

## References

- Clauson, G.  
1969 "Leksikostatičeskaja otsenka altajskoj teorii". *Voprosy Jazykoznanija*. 5: 22-51. Moscow.
- Dmitrijev, N.K.  
1955 "Čeredovanie glasnyx zadnego i perednogo rjada v odnom i tom že otdeľnyx tjurkskix jazykov." In N.K.Dmitrijev, ed. *Issledovanija po sravitel'noj grammatike Tjurkskix jazykov, čast' I: Fonetika*: 115.
- Doerfer, Gerhard.  
1985 *Mongolo-Tungusica*. Wiesbaden:Harrassowitz.

- Greenberg, J.H.  
1987a *Language in the Americas*. Stanford: Stanford University Press.  
1987b "Relative pronouns and P.I.E. word order type in the context of the Eurasiatic hypothesis", in Winfred Lehmann, ed., *Language Typology 1987, Systematic Balance in Language*,: 123-38. Amsterdam: Benjamins.
- Illič-Svityč, V.M.  
1971 *Opyt Sravnenija nostratičeskix jazykov*. Moscow: Nauka.
- Miller, Roy Andrew  
1991a "Genetic connections among Altaic languages", in Sydney M. Lamb and E. Douglas Mitchell, eds., *Sprung from Some Common Source*: 293-327. Stanford: Stanford University Press.  
1991b "Anti-Altaicists contra Altaicists". *Ural-Altaische Jahrbücher*. 63: 5-62.
- Poppe, Nikolaus  
1960 *Vergleichende Grammatik der altaischen Sprachen, Teil I: Lautlehre*. Wiesbaden: Otto Harrassowitz.  
1973 "Über einige Verbalstammbildungsaffixe in den altaischen Sprachen". *Orientalia Suecana*, 21: 119-41.
- Radloff, W.  
1882 *Phonetik der nördlichen Türksprachen*. Leipzig: Weigel.
- Ramstedt, G.J.  
1912 "Zur Verbalstammbildungslehre der mongolisch-türkischen Sprachen. *Journal de la Société Finno-Ougrienne*, 28.3: 1-86.
- Szemerényi, Oswald  
1970 *Einführung in die Vergleichende Sprachwissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Szinnyei, József  
1910 *Finnisch-ugrische Sprachwissenschaft*. Leipzig: Göschen.
- Tsintsius, V.I.  
1949 *Sravitel'naja Fonetika tunguso-man'čžurskix jazykov*. Leningrad: Učpedgiz.

## A Far-Out Equation

Eric P. Hamp  
*University of Chicago*

Our beloved and stimulating colleague Vitalij, whom we welcomed joyfully into our "neighborhood" nearly a quarter-century ago (a passage of time that scarcely seems possible) has devoted a third of a century and more to the relentless and noble search for perceptive and elusive glimpses at fleeting regularities that betray distant linguistic relationships; and he has insistently, correctly urged and exhorted us to join in such a search and to apply only the quality of criteria successfully evolved in the two-century search for nearer, more tractable relations. I agree entirely with Vitalij's basic principles, yet I must admit to having shown timidity in the past in considering certain specific problems of long-range relation. In that frame of mind I was certainly a rapid—and some would say hasty—doubter of the Nostratic hypothesis. I must always thank Vitalij (and a very few others, including Alexis Manaster Ramer) for having forced me to reconsider the theory and to reinspect the material with a more tolerant, liberal, and properly critical eye. I hope my future work will make this plain.

Of course my earlier work has had its bolder moments,<sup>1</sup> such as Maya-Chipaya; Eskimo-Aleut and Luoravetlan; Zuni and Penutian; Hurro-Urartian and Indo-Hittite. But it is also not correct to equate uncritical ignorance (if that is what it was) with boldness or courageous perspicacity.<sup>2</sup>

Some of my friends have also viewed my conclusions on certain closer relations as adventurous. Perhaps illustrations of this are to be found in Armenian *hariwr* = Greek ἀριθμός (KZ 72, 1955:244-45); Latin *mille*, *mīlia* (*Glotta* 46, 1968:275-77); Welsh *Mabinogi* = Latin festivals in *-ālia* and *-īlia* (see Paul Russell, *Celtic Word-Formation: the Velar Suffixes*, Dublin, 1990:60-61); Albanian *\*məl'ǎši* 'liver' (*Issues in Linguistics: Papers in Honor of Henry and Renée Kahane*, Urbana: University of Illinois Press, 1973:310-18).

In a vein that I hope may interest Vitalij, and that has always interested me, I want to dwell here on the value of a type of closer comparison that involves some of the analytic elements which become necessarily more common in the study of long-range relations. Perhaps we can learn something for our work in distant comparisons by considering attentively closer, more

---

<sup>1</sup> As well as its conservative ones, such as my reluctance to see Tovar's Celtic in the Indo-European of Spain (Iberia) until the stupendous and totally convincing discovery of Celtiberian in the Botorrita bronze; for a pre-Botorrita statement, see *Journal of Celtic Studies* 2, 1958:147-51.

<sup>2</sup> Cf. perhaps "Etruscan *max* '4'?" *Glotta* 37, 1958:311-12.

I have argued<sup>3</sup> that the derivation of Albanian *-zet* (*njëzët* '20', *dyzët* '40') '20' from *\*uik̥mti*: < *\*ui(:)k̥mti*: or *\*ui-dk̥mt-iH* < *\*dui-dk̥mt-i*? rests upon the recognition that *z-* here goes back to *\*uik̥-*.

Vedic *pāti-*: *patní-*; Greek πόσις: πότνια, δέσποινα; Latin *potis*, *-pte*

Skt. *viśpātiṣ*, *-ni*: ~ *viśāspātiṣ* (= Avestan *nəmo: vanta:*), *dāmpatiṣ*/δεσπότης, *kṣetrāsyapātiṣ*, *pātir dān*; Av. *šo:iθrapaitiṣ*, *de:ng paitiṣ* (*də:ng...*); Lith. *višpat(i)s*; = Old. Prussian *waispattin* (accusative), Lith. *-pati* f.

\**ueik*- [fem.] + genitive  $\Leftarrow$  + *póti*- [masc.] ['master']  
+ *pot(i)-n-iH<sub>o</sub>* [fem.]

> \**uik* -a:                    -s                    #*poti*- [masc.]; #*pot-ni-a*: [fem.]

> \**μik* -á:               #*p(o)ti-* [masc.];  
                                #*p(o)t-nja:* [fem.]

<sup>3</sup> See for earlier references my summary in Jadranka Gvozdanović ed., *Indo-European Numerals*, Berlin: Mouton de Gruyter 1992:900 and 919.

<sup>4</sup> That paper was not published at the time because I felt afterward that its content was not of primary interest to the field as it then appeared to emerge from the meeting, with a concentration on theoretical formalization that distracted from these data. A slightly inaccurate version of my paper and handout is reported by M. Huld, *Basic Albanian Erymologies*. Columbus: Slavica. 1984 s.v:137.

<sup>5</sup> I withheld that paper from publication because it conflicted seriously with a Hauptvortrag of the same meeting in 1978.

with loss of final *-s* (see especially *Romance Philology* 12, 1958:149-53) and with (compounding?) equivalent of Latin *-pte*.

> \**μiká:ti-* [masc.]; \**μiká:(t)nj-a:* [fem.]

with loss of stop before \**t* (\**nok<sup>w</sup>t-* > *natë* [fem.]; \**septṛṇ-ti-* [fem.] > *shta-të*)

> \**μká:ti-* [masc.]; \**μká:(t)nj-a:* [fem.]

with syncope of \**i* when unstressed (next to stress?) (\**di-t(i)-* > *ditë* 'day'; *so-t* ~ *so-d* 'today')

> \**μgá:ti-* [masc.]; \**μgá:(t)nj-a:* [fem.]

> \**gá:ti-* [masc.]; \**gá:(t)nj-a:* [fem.]

coarticulation

> \**ǰǰá:ti-* [masc.]; \**ǰǰá:(t)nj-a:* [fem.] (Roman period)

affrication of velar plus rounding plus palatalization; cf. Romanian.

> \**ǰot* [masc.]; \**ǰon-ə* [fem.] (arrival of South Slavs)

> \**zot* [masc.]; \**zon-ə* [fem.] (mid-first millennium CE)

> *zot*; *zonjë*, *zojë*

We will now use our experience with *-zet* and *zot* to tackle a more difficult problem which we find in Albanian *zog* 'bird'. We will find here that the necessary phonological sequence spans in a more crucial fashion the component morphological formants, thus imposing a far more demanding requirement for total accountability and full morpho-syntactic parsing—a requirement, of course, which in principle always applies. When it seems not to apply, we are simply lucky with our use of sloppy method.

Because of the rather unexpected correspondences in the equation which ensues, it may help to clarify matters if we first eliminate a clear Indo-European bird etymon, perhaps meaning 'eagle' or some large bird. On this etymon Greek gives us the most complex information: ὄρνις(:)ς, ὄρνις:θ- (but Doric ὄρνις:χ-) shows a stem \**ṛṇi:-* with ambiguous length (GEW 2.422 calls it a feminine -ι:-, but this really fails to explain adequately); however, accusative ὄρνιν etc. shows a clear *i*-stem of the type found in Greek, and ὄρνειον must be a thematization of this. In turn, as the Germanic evidence shows us, this *i*-stem (whose motivation, as I see it, would take us too far afield from our present argument) is affixed to an Indo-Hittite *n*-stem ὀρν-. From Germanic, ON *ari* shows us nominative singular \**are:n*, Gothic nom. pl. *arans* attests \**aran-ez*, ON



orn evidences accusative \*arn-un(z), while OHG *arn* and Old English *earn* 'erne' reflect oblique \*arn- beside German *Aar* from the old nom. sing. \*aro-. This *n*-stem, which we are led to by Greek and Germanic<sup>6</sup>, and which—we may note—shows a stable \*o vocalism in the base, is confirmed in an important way for us by cuneiform Hittite: nom. sg. *ḫa-a-ra-aš*, acc. sg. *ḫa-a-ra-na-an*, genitive sg. *ḫa-(a-)ra-na-aš* = ?/ḫarnas/. The reconstruction must therefore be with initial laryngeal (\* $\text{F}^w$  = H<sub>3</sub> or H<sub>0</sub>): \* $\text{F}^w$ Vro-, \* $\text{F}^w$ Vron-, weak stem \* $\text{F}^w$ Vrn-.

The Armenian *oror* 'seagull' agrees in the base, since *h*- would have been lost at the stage \* $\phi$  before a rounded *V*. Although I do not understand its nature, the same rhotic suffix seems to recur in Celtic: Old Irish *irar* 'eagle' > *ilar* (presumably by dissimilation) ~ *ilur* gl. *aquila* (*o*-stem masc.) < \**irora*- > \**irura*- < \**erura*- < \**eruro*-; apparently the thematization had (by IE rule) imposed \**e*-vocalism in the first position of the word, and therefore converted an earlier \**orur*- into \**eruro*-. The Welsh *eryr* 'eagle' (and comparable British forms), pl *eryr-od* with a suffix (old stem formant) found on animates and animal terms, can be either \**erir*- or \**orir*-, with umlaut; we might prefer the latter solution.

Finally, Slavic *орѣлъ*, Lithuanian dialect *arēlis*,<sup>7</sup> and Old Prussian \**arelis* point to \**or(e)l*-. The relation of \* $\text{F}^w$ Vr-l- to \* $\text{F}^w$ Vr-ur- or \* $\text{F}^w$ Vr-ir- is not clear to me. The actually attested Old Prussian *arelie* (V. Mažiulis, *Prūsų kalbos etimologijos žodynas*, I (A-H), Vilnius: Mokslas 1988:90) did not necessarily have an original older \**a*, as is shown by *addle* (Lithuanian *ēglė*), as 'ego' (Latvian *es*), *assanis* (Polish *jesień*), *assaran* (Latvian *ezers*, Czech *jezero*, Bulgarian *èzero*), as contrasted with *abse* (Latvian *apse*), *aglo* (=  $\alpha\chi\lambda\acute{\upsilon}\varsigma$ ), *ackis* (Latvian *acs*), *ape* (= Latin *amnis*), *assis* (Latvian *ass*, Latin *axis*), *awins* (= Homeric  $\alpha\omega\iota\varsigma$ ), *awis* (= Latin *auus* : Old Irish *áue* 'grandson' > Scottish Gàidhlic [o.ə] or [o.o]). But the sum total unobjectionally gives a base  $\text{F}^w$ Vr-.

If the base  $\text{F}^w$ er- 'large bird', suitably clarified, should prove illuminating for Nostratic, I would be very pleased. It will now be seen that this base does not underlie Alb. *zog*, pl. *zogh*, pl. definite (in archaic Tosk, especially Arvanitika, dialects) *zogh-të*. At this point it is useful to note the plural *zogh* [zɔj]; this must be derived from \*STEM-*i*: < \*STEM-*oi* (> north European IE \*-*ai*), a plural of a thematic *o*-stem. Thus, the plural guarantees a good age for the

<sup>6</sup> Alfred Bammesberger, *Die Morphologie des urgermanischen Nomens*, Heidelberg: Carl Winter 1990:176 concludes that we have \**ar-an*-.

<sup>7</sup> The Lithuanian *erēlis* with *e*-is one of a fair number of examples in Lithuanian where an initial *a*- has been taken as an \**o* and thus converted to a seeming ablaut \**e*-. The presence of Latvian *ereļi*: nominative plural and *ērglis* < \**erdlis* < \**erlis* (perceptively analyzed by Endzelīns) shows us that the \**e*- is real and of some age and therefore not a Lithuanian phonetic variant based on initial hard/soft phonotactics.

Albanian noun, surely pre-Roman; but the thematic stems were late and productive in IE, and probably rather recent in the prehistory of I-H. The last observation would make plausible the transparent morphonemic alternation of thematic *\*o ~ e* in relation to Auslaut.

We therefore have already recovered the stem form *\*-o-/e# > \*-a-/e#* of *zog*.

From the model of Albanian *\*ũĩk- > z-*, which we have already seen, our lexeme *zog* 'bird' easily suggests to us the noun seen in Vedic *vī-*; and as a matter of fact, in my 1969 paper I mentioned this and Avestan *vi-/vaya-* and the Nuristanī forms<sup>8</sup> without being able at the time to complete a plausible comparison and reconstruction. I hope now to repair that lacuna.

In the process of marshalling our data we will now briefly review relevant parts of the learned article by the late Jochem Schindler, *Die Sprache* 15, 1969:144-67 on "Die idg. Wörter für 'Vogel' und 'Ei'". For 'bird' Schindler adduces (146-47) the following lexemes: Vedic *vī-* masc., with nom. sg. *vīh ~ véh*, acc. sg. *vīm*, gen. sg. *véh*, nom. pl. *váyah*, instr. pl. *vīb<sup>h</sup>iḥ*, dat., abl. pl. *vībhyah*, gen. pl. *viná:m*; Avestan *vi-*, with nom. sg. *viš*, nom. pl. *vaiio:*, gen. pl. *vaiiām*, as well as the thematic *vaiiae:ibiia* and *vaiianām-ca*; Armenian *hav*, gen. sg. *havu*; Latin *avis* fem., gen. pl. *av-ium*; Umbrian acc. pl. *avif*, *avef*, *auif*, *aueif*, abl. pl. *avis*, *aves*, *aueis*.<sup>9</sup> I would add to this account that Armenian *hav*, with its *u*-stem, seen also in genitive plural *havuc*, offers no problem of stem class since (contrary to the reasoning of Schindler, 158 §4.4), in Armenian the *u*-stems were productive for animal names, and it seems that what the *\*-u*-replaced was taken by speakers to be an *i*-stem.

Greek furnishes us with *ἀλῆτος ~ ἀλῆτος* 'eagle', and *αἰβῆτος*, which Hesychios attributes to the *Περγαῖου*. These have all been traced to *αφῶ-ετος*, with the morphology of *νυφετος*. We thus find here *\*aui-*, which is ambiguously *\*H<sub>a</sub>eui-* or *\*H<sub>a</sub>ui-*.

Schindler's account (p. 147) of Welsh *hwyad*, Breton *houad* 'duck' is inconclusive, but I consider these to be unrelated. F. O. Lindeman's doubts

<sup>8</sup> I had been working intensively at times in 1965-68 on Nuristanī, but the Russian invasion of Afghanistan later put an end to my hopes—hopes that I have never abandoned.

<sup>9</sup> Schindler reports Szemerényi as finding *\*aui-* in OCS *žeravъ*, *γέρανος*, *grūs*; in passing, we might remark that Russian *žurávl'* must come from Polish *uraw = óraw*. Now surely, as Machek implies s.v. *jeřáb ~ řěáb < žeřáb < žeřáv* in Czech, this bird name must be somehow connected with *γέρανος*, *grūs*; *crane*, etc. Vasmer, *Russisches etymologisches Wörterbuch*, 1:433-34, offers no explanation for the comparisons given, but the motivation for such a compound, instead of suffix, then vanishes. Indeed, we seem to have a suffix alternation *\*u ~ n-*, perhaps an old heteroclitite.

(*Bulletin of the Board of Celtic Studies* 30, 1983:303-04) based on a dialect variant in Welsh are unfounded, I believe, and fail to recognize a secondary *chw*-. For *hwyad* = *houad* I follow Lockwood, and I have proposed (*Zeitschrift für celtische Philologie* 43, 1989:196-97) *\*sisat* < *\*ses-a-t* replacing older *\*an-a-t*-. Thus we conserve the old IE 'duck' etymon. A conservation is always to be preferred.

Schindler's treatment of Albanian *vito*, *vido*, North Geg *vid(ë)*, Arbëresh *vidhez(ë)*, based on Jokl, cannot be sustained, and these will be analyzed below.

Clearly correct is Schindler's statement (147), "Die Untersuchung muß sich im wesentlichen auf ar. *\*vi*- und lat. *avis* stützen." However, I would now modify this 1969 statement to take proper account of Arm. *hav*, which shows us the reflex of the specific initial laryngeal (pace Schindler). We therefore reconstruct Lat. *avis*, Arm. *hav* < *\*ǵVui*-, Vedic *vi*-, Avestan *vi*- < *\*ǵui*-. Our reconstruction corresponds in a modern way to Pokorny's (*IEW*, 86) *\*aǵei*, *\*əǵei*; I am, however, not sure that Avestan really belongs here also with *a:-vayeiti* 'fliegt heran' (which verb of motion does this really continue?). Therefore we will not discuss here a verbal base *ǵuei*-.

Schindler then embarks on a long discussion (148-59) of the possible solutions to be offered for the IE structure and later dialect development of the underlying root (and other perhaps parallel or analog IE root types) and nominal paradigm of our noun. This rich and suggestive discussion is offered as an approach to a solution because the evidence for the attested descendant nouns is too ambiguous to yield a direct formulation of the parent IE paradigm. However, the discussion and critique which this valuable contribution of our lamented colleague and friend Schindler to our understanding of IE noun inflection merits today would take us too far afield from our proper subject and would require space we do not have. Furthermore, the precise form of the strong-case stem of our noun would not affect the reconstruction with which we are here occupied.

We are also relieved of the need to consider the difficult Hittite *šu-wa-ış* = /swais/ 'Vogel' < *\*suojs* (?) adduced by Schindler 159 §5. This word will not affect our Albanian reconstruction.

Schindler's conclusion (167 §10), regardless of whether we accept his stem shape *\*h<sub>2</sub>u<sub>2</sub>oj-* ~ *\*h<sub>2</sub>u<sub>2</sub>ej*, that this noun 'Vogel' was a Wurzelnomen certainly remains correct. One aspect, however, that of the Vedic accent, remains outside the reckoning of Schindler. A. A. MacDonell, *A Vedic Grammar for Students* (Oxford 1916:458, §c.1 note) observes that the accent of *ví-*, contrary to that of normal monosyllabic (i.e. oxytone) stems, fails to move to the ultima in weak cases: *ví-bhiṣ*, *ví-bhyas*, but *vi:-ná:m* (surely a renewed and contaminated formation; cf. Avestan *vaiiam*). This behavior must point to an old disyllabic base for this root noun; note the same pattern in (*H<sub>a</sub>*)*nṛ* 'man', *kṣám* 'earth', *svár*

'light', *śván* 'dog', *yúvan* 'young', all of which I have discussed elsewhere for aspects of their disyllabicity.

I therefore reconstruct schematically  $*\text{ǵVuéi-} > *\text{ǵV}\text{ǵi-} > *\text{Ha}\text{ǵi-} >$  Latin *avis*, Umbrian *avi-f*, Armenian *hav*;  $> *\text{ǵyéi-} >$  Vedic *vé:-h*, *váy-aḥ*, Av. *vaiio*;; weak-state stem  $*\text{ǵuí-} >$  Indo-Iran. *ví-*, Greek *αἰβετός*, *αἰετός*. We are now prepared to turn to the Albanian forms.

We must first address *vido*, *vito* 'dove', which N. Jokl attempted to explain, but surely without success, in *Linguistisch-kulturhistorische Untersuchungen aus dem Bereiche des Albanischen*, Berlin, Leipzig: de Gruyter, 1923: 299-301. Likewise unsuccessful was Meyer's appeal to an animal call rejected by Jokl and revived by E. Çabej, *Studime etimologjike në fushë të shqipës*, vol. 1, Tiranë 1982: 72, 130 (French 196, 287). Jokl's reconstruction would be IE 'bird' + collective suffix  $*-d-$ , but this is too vague as well as phonetically imprecise and morphologically incomplete. Jokl fails (p. 300) to specify the vowel grade of 'bird', merely citing the Vedic and Latin nouns; it will be seen that I assume for this part of the formation a stem  $*\text{ǵyéi-}$ . Jokl claims that the collective suffix (without justifying the semantics of *collective* here),  $-d-$ , was devoiced in final position to  $-t$ . That is of course quite possible, but fails to specify the conditions of Auslaut, to explain why we should have *both* reflexes in  $-d-$  and  $-t-$ , and to clarify why final should now be medial; such an explanation is totally *ad hoc*. I have pointed out that a post-tonic vowel<sup>10</sup> between two \*dentals, probably of different voicing in the order voiced—voiceless, was syncopated, resulting in medial duplicate (dialectal) reflexes  $d \sim t$ ; see *Acta Linguistica Hafniensia* 12, 1969: 154, *Glotta* 50, 1972:299, and *Revue roumaine de linguistique* 18, 1973:337. In this way I believe I have opened the path to an adequate explanation of *sot*  $\sim$  *sod* 'today', *vitol/vitua*  $\sim$  *vidë/vidâ* 'dove', *gat/gāti*  $\sim$  *gādi* 'ready, prepared, timely, almost'  $< *\text{g}^{(h)}\text{ād}^{(h)}\text{it-}$  ( $\rightarrow$  Romanian *gat-a*  $<$  *gat* + *illac*), *lot*  $\sim$  *lod* 'tears (wept)'  $< *(s)\text{leiḡ}^{(h)}\text{-V-to-}$  ( $>$  pre-Roman period  $*\text{lé:dōVto-}$ ); and *lodrë* 'game'  $< *\text{leid-V-tra-}$  ( $>$  pre-Roman  $*\text{le:dVtra-}$ ) beside *lojë*  $< *\text{leid-rja-}$ . I have explained the development of  $*ei$  after  $*l$  to  $*e$ :  $> o$  (which was not yet explained in *Glotta* 50, 1972:299) in *Gjurmime Albanologjike* 6, 1978:41-42 (Prishtinë). We thus reach a set of reconstructions:

*vito/vitua*<sup>11</sup> fem. (strongly Tosk)  $< *\text{ǵyéi}(s) + d^{(h)}\text{Vté:}$  or  $*\text{ǵyí-d}^{(h)}\text{Vtè:}$   $< *\text{ǵ}^{h}\text{uptè:}$   $< *\text{ǵ}^{h}\text{ub}^{h}\text{-te:(n)}$ .

<sup>10</sup> At first I specified  $*i$ , but I have since seen that the quality range must be viewed as broader, though probably a high vowel.

<sup>11</sup> *vido*, cited by G. Meyer from Mitkos (19th century), is not shown in dictionaries, but apart from Pokorny *IEW*: 86, lives on in Jokl, 1923 and Schindler op. cit., 1969. If *vido* should be a *vox nihili* its absence will not damage our analysis.

*vidâ* masc. (Geg) < \*ǵuēi(s) + d<sup>(h)</sup>Vtán- or \*ǵuī-d<sup>(h)</sup>Vtàn- < \*±d<sup>h</sup>uptàn- < ±d<sup>h</sup>ub<sup>h</sup>-ton-

*vidë* fem. (Geg, Arbëresh Tosk) < \*ǵuēi(s) + d<sup>(h)</sup>Vtá: or \*ǵuī-d<sup>(h)</sup>Vt(-à:) < \*±d<sup>h</sup>upt-à: < \*±d<sup>h</sup>ub<sup>h</sup>-ta:

In these forms we must recognize, besides the IE etymon of 'bird' and the possibility that we have an ancient phrase (\*'bird, i.e. dove') or perhaps a compound (\*'a dove-bird'), different but related suffixal derivations of the IE base for 'dove, Taube' \*IEW: 284).<sup>12</sup> It is *dut-*, or perhaps *-dut-*, which gives *d ~ t*.

Jokl also gives, and Schindler following him, Arbëresh *vidhez(ë)*, i.e. with hypocoristic (fem.) -zë; this form was taken from Guiseppe Schirò (1865-1927), who was a native of Piana degli Albanesi, Sicily, and not from a dictionary. We now have a dictionary, though not yet a complete survey of Arbëresh, Emanuele Giordano, *Fjalor i arbëreshvet t'Italisë*, Bari: Paoline 1963; on p. 542 we find: *vidhëz* (Schirò, and from Pallagorio in the Crotone area of Calabria), *vidë* (Schirò) and *vide* (also from Sicily). This is a very poor and meager sample, and we badly need at least another thirty well-chosen examples (and then these multiplied by some factor to avoid error in collection and recording). But, as an attempt at interpretation for the present, I suggest that *vidhëz* reflects a conflation of *vidhe-z* with *vidë*; these speakers easily distinguish *d* and *dh* [ð]. Then *vide* confirms the reality of *vidhe-z*.

Thus we see that *vidhë-z* may well not be original. Similarly, *vide* is hard to explain;<sup>13</sup> it could be latterly patterned on *vidhe-z*. We have already explained above *vidë*, which is supported by Geg. We now have only *vidhe-zë* to explain.

If we start from a phrase of the type *blackbird*, *bluebird*, *goldfish*, *whitefish* (as opposed to the nominalized derivatives *grayling*, *the bear* < \**ber-an-*, presumably 'brown one') we may construct forwards:

\*ǵuēis + d<sup>h</sup>eu(H)b<sup>h</sup>-a: 'dark (?) bird, i.e. dove' > \*uēi + déuba: > \*uēi + ðéuba: > \*uī:(+)ðéuba > uī:ðé:Λ > \*viðé (+ \*-dja: > -zΛ) *vidhe-zë*.

It will be seen here that it is the quality of the modern *e* vowel which (by reasoning of elimination) leads us to the \**eu* diphthong which in turn points

<sup>12</sup> Pokorny's account, pp. 283 ff. is one of his worst. There are clearly different original bases involved here, 'smoke, dust; dark, black; deaf, dumb', a derived sense for the dove, but surely not water in any direct way. Perhaps there was also a root *d<sup>h</sup>euHb<sup>h</sup>*.

<sup>13</sup> If *-d-* is from *-dVt-*, *-e*, which is usually from *-ja:*, would not be expected here following a dental, since normally we expect *\*-tja:* > *-së* and *-dja:* > *-zë*. Of course *\*-dVt-* > *\*-dd-* could complicate the issue.

to the state  $*d^h eu(H)b^h$ - of the base and the consequent morphology and syntax that is likely; this also gives us the possibility of losing by rule the medial voiced obstruent  $*b < *b^h$  (cf. *ve* 'widow'  $< *medV\mu\Lambda < *mid^hV\mu a$ ; Sofikò Arvanitika *káalə*, pl. *kúaλ* 'horse'  $<$  Latin *caballus*, *caballī*). We have now accounted for all of the phonology, including the prosodics, while hypothesizing plausible morphology and syntax, and identifying lexical elements with their semantics. A maximum of parsimony (= continuity) has been observed in the mapping on known IE and relevant IE-dialect elements. The presence of the IE etymon  $*\xi Vu\acute{e}i$ - in the Albanian lexicon is assured.

It is time now to consider *zog* 'bird.' Clearly the most parsimonious solution will be to find  $*\xi Vu\acute{e}i$ - in this lexeme. This means finding the descendent of  $*\xi V\acute{u}ei$ - in *z*-, as was suggested above when we first turned to the features displayed by *zog*. To do this we will require  $*\xi V\acute{u}ei$ - in the form  $*\xi \mu i$ -  $> * \mu i$ -, which clearly draws upon the original combining form of the base, as opposed to *vidhè-zë* above.

The etymology of *zog* has a very sparse history. G. Meyer in his 1891 etymological dictionary had two tentative suggestions: One links *zog* with *zot* and tried to construct a derivation from IE  $*\hat{g}en?$ - (in modern terms) 'procreate, be born'; apart from the vocalism, we have known since Pedersen in 1900 that this will not work, since  $*\hat{g}$  will not produce *z* in this environment. The other suggestion simply mentioned Sanskrit *jahu*- 'young animal' without pressing the matter further. With Grassmann's dissimilation this could lead to  $*\hat{g}^h V\hat{g}^h$ -, but not to  $*\hat{g}^h \mu V\hat{g}^h$ - (to accommodate *z*-), or  $*g^{w(h)}e\hat{g}^h$ - (for *z*- by Pedersen's discovery of 1900), but lacking an Indic context to palatalize a root-final  $*g^{w(h)}$  for *zog*. Then in 1892 (*Albanesische Studien* III:18) Meyer adduced (from reading Lagarde) Armenian *jag* 'junger Vogel; also young animal' and reconstructed  $*\hat{g}^h a(:)g^h(u)$ -.<sup>14</sup> Disregarding the semantic requirement of 'young', we now know that only  $*\hat{g}^h \mu Vg^{(w)h}$ - will suffice (but not for the Indic) if Armenian *jain* 'voice' really matches Albanian *zâ*.<sup>15</sup> Pedersen and Walde-Pokorny later supported this equation (with  $*-g^{wh}$ ). Tagliavini (1937), who depended heavily on Jokl, had nothing to add to this, ending his account, overly brief and unexplanatory in itself, with the citation of  $*\hat{g}^h a:g^{wh}$ -, which will of course *not* explain the *z*-. Çabej, *Studime gjwhësore*, II Prishtinë: Rilindja 1976:327-28, gets no further than this (nor does Huld, 1984), other than to suggest additionally the impossible Lithuanian *jegà* 'power' and its etymon, as supporting cognacy for the sense 'young.'

<sup>14</sup> G. B. J&ahukian, *Sravnitel'naja grammatika armjanskogo jazyka*, Erevan: ANASSR, 1982:49 still supports *jag*  $< * \hat{g}^h a:g^{uh}$ -.

<sup>15</sup> J&ahukian op. cit. (1982:75) recognizes *jayn*  $< * \hat{g}^h \mu \eta ji$ -, comparing Slavic *zvonъ*, the cognate of Albanian *zâ*.

The only increment to this scholarship known to me is the study by Bojan Čop, *Živa Antika* 3, 1953: 17 (page number on offprint), attempting to relate Greek  $\phi\acute{\alpha}\psi$ ,  $\phi\alpha\beta\acute{\omicron}\varsigma$  and  $\phi\acute{\alpha}\sigma\sigma\alpha/\phi\acute{\alpha}\tau\tau\alpha$  'wild pigeon' to *zog* through the reconstruction  $*\hat{g}^h\mu a(:)g^{wh}$ . We see that this is a very intelligent reconstruction, fulfilling the requirements specified above, as none of the others does. The one way we could modernize this reconstruction is by specifying an old ablauting root noun,  $*\hat{g}^h\mu(e)Hag^{wh}$ . However, it must be acknowledged that there is a difficulty in getting from  $\phi\alpha\beta$ - to  $\phi\acute{\alpha}\sigma\sigma\alpha/\phi\acute{\alpha}\tau\tau\alpha$  by iotization, since  $\beta\acute{\iota}$  (i.e. as a labio-velar) should yield  $-\zeta$ -. Frisk (*GEW* 2:996-97) duly notes this, and calls  $\phi\acute{\alpha}\psi$  and  $\phi\acute{\alpha}\sigma\sigma\alpha$  unexplained. Nevertheless, there is a neutralization argument which could be correct in justifying Čop's equation. With Čop's argument, then, we have a possible equation for Greek, Armenian, and Albanian.

But the weakness in this etymology is the complex, ambiguous, and otherwise unmotivated root shape, the admission of deviant Greek consonantism, the fact that the Armenian initial really does not require correspondence, the subgrouping relation of Greek and Armenian, which subtracts from the independence of their testimony, and, especially, the marginality of the semantics as a correspondence in the Greek and Armenian. For these reasons we have been encouraged to seek a solution elsewhere.

We will first complete the account of *z-* in *zog* by adding to  $*\xi\mu i$ - the pretonic  $*\hat{k}$ , which we have seen in *-zet* and *zot* above. In a so-called kental language  $*\hat{k}$  would of course appear as  $*k$ . If we consider animal names, we find that diminutives ( $\rightarrow$  hypocoristics) are moderately frequent. The detail of diminutives in the various IE branches is varied, multifarious, often renewed, even unequal in its density or frequency: i.e. in some branches diminutives are much more favored and invoked, and as a result become bleached and lexicalized. Thus, e.g., what we find rare or highly marked in Classical Latin or Greek may become the basic colorless term in the descendent languages. We will look at Latin, which, as an Italic language is a European IE language relevant to the (North European) language that is Albanian. Latin had a well-known diminutive formation<sup>16</sup> with a thematic suffix in  $*-l$ -:

<i>agnus</i> 'lamb' :	<i>agne-llus</i>
<i>haedus</i> 'kid' :	<i>haedi-llus</i>
<i>uitulus</i> 'calf' :	<i>uitel-lus</i>
<i>canis</i> $\rightarrow$ $*catos$ :	<i>catu-lus</i> $\rightarrow$ <i>catel-lus</i> 'puppy'
<i>porcus</i> 'pig, hog' :	<i>porcu-lus</i> $\rightarrow$ <i>porcel-lus</i>

<sup>16</sup> I have argued elsewhere that not all suffixal specimens of the same shape have the same value, e.g. *oculus*; *famulus*, *bibulus*.

In folk Latin the noun *auis* was used with a velar suffix:

*auis* → \**auica* > \**auca* > Italian *oca* 'goose'; French *rue aux Oues* > *oie* 'goose'

We find that the etymon of *ouis* was used in IE with a velar suffix:

*ouis* 'sheep' → *ouicu-la* → \**ouicel-la* > Spanish *oveja*

And the same pattern is repeated for *auis*:

*auis* : *auicu-la* → \**au(i)cel-lus* > French *oisel*, fem. *oiselle* (1562 Rabelais) → *oisel-et* > *oiseau*

We see, then, that \**ǵVu(e)i*, perhaps when suffixed<sup>17</sup>, becomes \**ǵVu(e)i-k̑*. Hence, thematized with no case inflection: \**ǵui-k̑-e*, equivalent to \**auica*; this \**ǵuike-* is nearly identical phonologically with *auicu-la* (: \**au(i)cel-lus*) < \**ǵVuike-l-a*; the stem in the Latin diminutive. An unrecognized cognate of this formation is found in Lithuanian *vīšt-à* 'hen, chicken' < \**ǵuik̑-t*.<sup>18</sup>

We must now identify a comparandum to complete the -*g* of *zog*. I suggest that the closest analog is to be found in Sanskrit *patam-gá-* adj. 'flying'; masc. 'bird.' Other related formations are Sanskrit *vanar-gú-* 'wandering in the forest', Lithuanian *žmo-gùs* 'man', Armenian *bok* 'barefoot' < \**b<sup>h</sup>oso-g<sup>w</sup>o-*; see E. P. Hamp, "Armenian *bok* 'barefoot'" *Revue des études arméniennes* (n. s.) 20, 1986-87: 35-36. Here we have a quasi-compound with a second element that in IE was in the process of transferring from the function of base to that of a highly specified suffix. This suffixal, rather than compounding, status will be seen to be confirmed by the vocalism \**e* preceding the *g* < \**g<sup>w</sup>* in *zog*; that is, the vowel here was not the IE compounding \**-o-*, which would have given *-e-*, i.e. †*zeg*.

Notice that the semantics of *zog* are seen to be those of Armenian *bok* and Lithuanian *žmo-gùs*; that is to say, \**b<sup>h</sup>oso-g<sup>w</sup>o-* ≡ \**b<sup>h</sup>oso-*, and *žmo-gu-* ≡ Old Lithuanian *žmuõ*, Old Prussian *smoy*. In other words, our ex-compound has the same value semantically as its simplex, or base; and the incipient suffix marks it as animate. In Sanskrit *vanar-gú-* and *patam-gá-* we may see more living predicate complements to *-gulga-*.

<sup>17</sup> \**auic-a* is as though thematized.

<sup>18</sup> I ventured this tentatively in *Baltistica* 3, 1, 1967: 8.



Unifying our elements, we have now reached for *zog* a reconstruction \**ʕui-k-e-g<sup>w</sup>-o-* 'bird + stem-formant (with suffix)<sup>19</sup> + *thema* + going (animate) + *thema*'. At the beginning of our discussion we recognized that the stem of *zog* was thematic; hence the final *thema*.

We now recall that, as I have stated elsewhere, Albanian underwent the same lengthening of syllabics in position before IE mediae that Werner Winter has discovered for Balto-Slavic. Therefore:

\**ʕuike(+)g<sup>w</sup>-o-* > Alb.-BS \**ʕiuke:g<sup>w</sup>a-*. This leads us to Albanian \**ʕiké:g<sup>w</sup>a-s* > \**ʕġé:g<sup>w</sup>a-* > Roman period \**ʕġō:ga-* > Slavic arrival \**ʕġog* [masc.] > mid first millennium CE \**zog*, plural \**zōg*.

It seems to me that there can be no doubt as to the outcome of \**ʕui-k-e(+)g<sup>w</sup>-o-* in Albanian *zog*, both meaning 'bird (animate)'. We see here an intra-familial comparison with the involvement of radical phonetic change over time and a moderately complex syntactic morphology, and therefore with the absolutely essential requirement for an exact and exhaustive accountability for all elements and features. Of course, this requirement imposes a careful standard of control on relevant chronologies.

It frequently becomes necessary for us to correct or refine past scholarship along the way, including our own. We must also expect our own work to be refined in the future. Present efforts rarely reach ultimate answers. But we must work as if they did.

---

<sup>19</sup> Visible in Latin, Balto-Slavic, and Albanian; I will deal elsewhere with Slovene dialect *wtíc*, Slovak *vták*, which are unrecognized beside Russian *ptíca*, Serbo-Croatian *pàtka*, Czech *pták*, etc.

# On Grammaticalization in Nostratic

Irén Hegedűs

*Janus Pannonius University, Pécs, Hungary*

## 1. Preliminary notes on grammaticalization theory.

This paper discusses a case of grammaticalization, using the example of reconstructed Nostratic morphemes. It intends to demonstrate that the Proto-Nostratic locative particle is derivable from the PN lexeme meaning 'nearby.' The cycle of grammaticalization seems to be complete in those Nostratic daughter languages where the PN locative particle is reflected as a bound morpheme. In this respect Afroasiatic stands apart from the rest of the Nostratic family because the cycle stopped at the level of cliticization.

The rich literature on the theoretical implications and ramifications of grammaticalization is not our main concern here. In order to shed some light on the theoretical background of my argumentation, I would make a brief note on my understanding of the notion of grammaticalization. In this respect I follow the ideas of Jerzy Kuryłowicz, just as some leading experts in current grammaticalization studies do (cf. Heine et al. 1991: 4, 24). The diachronic process in which a content word (lexeme) can acquire the grammatical characteristics of a function word and becomes a clitic or an affix will be called grammaticalization. This definition of mine is more or less congruent with the notion of grammaticalization applied by Heine et al. (1991) and Hopper and Traugott (1993). Furthermore, I consider the transformation of a lexical item (free morpheme) into an affix (bound morpheme) functioning as a grammatical marker to be the most complete instance of grammaticalization. Not all grammaticalization processes are necessarily carried out to a stage of completion: the diachronic development may stop at some stage along the cline of grammaticality established by Hopper and Traugott as follows:

content item > grammatical word > clitic > affix (cf. Hopper and Traugott 1993:7).

## 2. An earlier suggestion concerning instances of grammaticalization in Nostratic.

In an article published in 1971 (and republished in English in 1992) Aron Dolgopolsky suggested that in two cases Nostratic free morphemes could have been the sources of bound morphemes (cf. Dolgopolsky 1971, 1992).

One of these cases is the PN lexeme \**na*<sup>c</sup> *a* 'to go to do something'. The reconstruction of this PN etymon was based on evidence from IE, AA, K and A (Tungusic); from this protollexeme Dolgopolsky derives various

(dominantly modal) affixes in the Nostratic daughter languages such as:

AA: Old Assyrian conjunctive *-ni*, Arabic modus energicus *-n*, *-nna*, Cushitic (Bilin, Kemant) optative-jussive *-in*, etc.

K: Old Georgian 3rd pers. imperative *-n*, Swan 'zaglaznoe naklonenie' *-un-i*, *-əni*, *-in-i*;

IE: present (imperfective) *\*-n-*;

U: Finnish potentialis *-ne-*, Hungarian conditional *-na-/ -ne-*, Mansi conditional *-nuw-*, Selkup conjunctive *-ni-*, *-ne-*, etc.;

A: Tungusic *\*-na-/ -nä:-* suffix meaning 'to go to do something'.

Since Dolgopolsky does not provide a reconstructed PN form of the grammaticalized affix, it seems that he supposed the grammaticalization to have occurred in the individual N dialects. Furthermore, it is not clear whether Dolgopolsky has in mind the same morpheme as the PN medio-reflexive *\*-nV* mentioned but not analysed among verbal voice suffixes in the Introduction to the Nostratic Dictionary (Illič-Svityč 1971:13). There has been no comment on this morpheme in the later literature.

The other case of grammaticalization suggested by Dolgopolsky is the development of the PN lexeme *\*šew<sup>c</sup>V* 'to want, to agree, to allow (> to ask)'. This lexeme was grammaticalized as a causative-desiderative affix (cf. Dolgopolsky 1971:240-242 or 1992:293-95). It is again not clear if Dolgopolsky's suffix is supposed to be identical with the causative-desiderative suffix (*\*-sV*) which is mentioned in the Nostratic dictionary in the tables of the Introduction but does not constitute an entry in the dictionary (Illič-Svityč 1971:13). The case for this morpheme may be stronger than for the previous one, although Alexis Manaster Ramer, discussing the problems encountered in the semantic reconstruction of Nostratic etyma, has brought up a serious counterargument: "Even less convincing are many of the comparisons involving affixes, such as the proposed Nostratic 'causative-desiderative' *\*-sV*, whose reflexes have desiderative senses in Indo-European and Altaic and causative senses in Dravidian and Afroasiatic (Illič-Svityč 1971:13), so that there is no evidence linking the two" (Manaster Ramer 1993:223). It is indeed true that the distribution of semantic features is uneven.

There might be, however, a feasible explanation to this. Dolgopolsky (1971:241; 1992:294) gave the following scenario for the emergence of this Nostratic morpheme. A PN free morpheme *\*šew<sup>c</sup>V* 'to want, to agree, to allow (→ to ask)' is reconstructable on the basis of IE, AA, K, D, A (Tungusic) and somewhat ambiguous U data. (N.B. Illič-Svityč (1967:358) reconstructed *\*šVwV* on the basis of the same IE and K data that Dolgopolsky used.) This lexeme - according to Dolgopolsky - frequently occurred in the analytic construction {X + *\*šew<sup>c</sup>V*} 'to want X, to agree to have X', from which the seemingly irreconcilable meanings of causative and desiderative can perhaps be

derived. Moreover, this grammaticalization may have occurred in the era of disintegrating PN. Thus it could have yielded a morpheme with a desiderative meaning in some daughter languages (IE and A) and a morpheme with a causative meaning in other daughter languages (D and AA). So the functional development was divergent already in the process of grammaticalization. This scenario, however, may not necessarily disperse all doubts, and beside the ambiguous semantic development the phonological side also needs clarification (initial *š* in the free morpheme becomes intervocalic *s* in the affix?).

### 3. PN *\*daḲa* 'nearby' > PN *\*da* 'locative particle' > locative-ablative suffix.

Locative suffixes can often be derived from earlier postpositions that themselves may derive from earlier nouns (cf. Hopper-Traugott 1993:107-108). Below, the case of a PN locative morpheme will be examined as a possible derivation from a PN content word.

#### 3.1. Stage 1: content word.

The PN free morpheme *\*daḲa* 'nearby' was posited on the basis of Uralic, Altaic and Afroasiatic evidence (cf. Illič-Svityč 1971: 215; Nr.61), and this reconstruction was later supplemented by an IE (Hittite) reflex proposed by Václav Blažek (1989). Since only a single lexeme represents the IE family, the IE evidence is too weak in my opinion. A brief summary of the evidence can be given as the following:

1. PU *\*taka* 'rear'; *\*taka-na* 'behind' (cf. UEW, p.506-507: PU *\*taka* 'Hinterraum, das Hintere') with reflexes both in the Finno-Ugric and in the Samoyedic branches; in the U daughter languages it survives as a noun and it also serves as a base for several postpositions in Lappish (e.g. *tuoke:n* 'behind, beyond', *tuoke:s* 'von einem Platz hinter etw.; hinter etw. hervor, heraus', etc.) (UEW, *ibid.*).
2. PA *\*daka-/daga-* 'near, to near, to follow somebody' (this PA etymon occurs as *\*daḡa* 'to follow, accompany' > PMTung. *\*daga-* 'near': Tung. *daga* 'near' in UEW, p.507).
3. PAA *\*dk* 'nearby' (Cushitic and Chadic reflexes).
4. PIE ? : Hittite *taki-* 'other' (Blažek 1989).

The capital *Ḳ* in the reconstructed Nostratic form is either a glottalized velar stop *ḳ* or a glottalized postvelar stop *q*. Since the reflexes of these Nostratic phonemes merged in all the descendants except for Kartvelian, we could only establish the exact nature of the second consonant in the Nostratic stem if we had a Kartvelian reflex, which - to the best of my knowledge - so far has not been

found.

### 3.2. Stage 2: content word grammaticalized as a particle.

The Nostratic lexeme *\*daḲa* must have undergone a process of attrition, i.e. a gradual phonological erosion and semantic fading that lead to an abstract notion. The circumstance that can facilitate such a process of lexeme attrition is the increasing frequency of use; this contributes to the growth of functional load compensating for the semantic bleaching of a lexeme. It seems to be phonologically and semantically feasible that the PN locative particle *\*da* could have emerged in a process of grammaticalization from the PN free morpheme *\*daḲa* 'nearby'.

### 3.3. Stage 3: particle grammaticalized as a suffix.

Illič-Svityč reconstructed a PN locative particle *\*da* (cf. Illič-Svityč 1971: 212-214; Nr.59). The following reflexes of this particle show that, in the further process of grammaticalization, PN *\*da* yielded a clitic in AA, while in other Nostratic daughter languages the process resulted in affixes:

1. PAA *\*d* 'particle with locative meaning'; reflexes in Berber as directional clitic with verbs, in Cushitic as postpositive locative particle (both verbal and nominal). The PN particle is preserved as a free morpheme only in this branch; in the other Nostratic branches it appears as a bound morpheme. The AA family most probably has the largest time-depth and therefore can be considered the earliest diverging Nostratic branch. The fact that only AA has preserved the PN locative particle as a clitic and not just as an affix also supports the conclusion that here we are dealing with an archaic N feature in AA and that this might be another instance showing that AA is very close to the PN stage.

2. PK *\*-da* 'directional-locative suffix (of pronouns and adverbs)' and PK *\*-d/-ad* 'dative suffix of nouns' are listed as evidence in the Nostratic dictionary (cf. Illič-Svityč 1971: 212-213). These Kartvelian data are somewhat problematic for two reasons. First of all, the affix *-da* can be found on personal pronouns in Georgian, Megrel and Chan only, so Klimov is obviously correct to think that *\*-da* can be reconstructed for the chronological level of the Georgian-Zan unity (cf. Klimov 1964:43). Thus, the Kartvelian evidence is chronologically rather shallow, unless we hypothesize that the morpheme was lost in the rest of the Kartvelian family.

The other reason that makes the Kartvelian evidence problematic is the strong functional divergence, i.e. it is mostly a directional or dative meaning that is carried by the Kartvelian morphemes. There seems to be no trace of the ablative.

3. PIE *\*-ed* 'ablative suffix in personal pronouns' (Beekes 1995: 208), *\*-o:d* 'ablative suffix in *o*-stems and in masc. sing. indefinite pronouns' (Beekes 1995: 204; Szemerényi 1990: 197, 219). PIE *\*-o:d* can derive from *\*o* + *\*-ed* < *\*oh<sub>1</sub>ed* (Beekes 1995:192). The Nostratic dictionary has the PIE form as *\*-D/-eD* 'ablative suffix in pronominal and *o*-stems' (Illič-Svityč 1971: 212), where the capital *D* stands for *t/d*. In the 1960s it was general to posit an archophoneme in the ablative suffix (cf. Kazanskij 1989:123), and this assumption is reflected in the Nostratic dictionary. Since then not much specification has been achieved in the reconstruction of this morpheme. From the aspect of Nostratic phonological correspondences we would expect an aspirated voiced stop in the PIE ablative suffix (PN *\*d* > PIE *\*dh*), but there are certain indications in the literature of the 1980s that the PIE dental ablative suffix may have contained an aspirated voiced stop (cf. Cohen 1984, Shields 1987). Kenneth Shields derives the dental ablative suffix from a deictic particle in *\*-dh* and supposes that the ablative suffixes with unaspirated dentals emerged as sandhi variants (Shields 1987:63, 66). Thus it seems to be feasible that the PIE ablative suffix emerged from an earlier deictic particle *\*-dh* that can be connected with the PN locative particle *\*da*, grammaticalized from PN *\*daḲa*.

4. PU *\*-δa/-δä* 'ablative suffix (in pronominal and averbial stems)' (Illič-Svityč 1971:212-13, Raun 1988:559). The status of *\*δ* in the PU phonemic inventory has been the subject of numerous studies (for a summary see Honti 1992). Its reconstruction seems to be indispensable because - in its absence - certain etymologies would remain unexplained (cf. Lakó 1968:68-69, Itkonen 1969). This phonological segment is still used in Rédei's UEW but it never occurs in initial position. Its palatal pair *\*δ'* appears only in four etyma, of which only two (*\*δ'eme* 'Traubenkirsche, Ahlkirsche' and *\*δ'imä/δ'ümä* 'Leim') are securely reconstructed for PU (cf. UEW, pp.65-66). This deficient distribution may imply that in early PU they were still allophones; especially since PU *\*δ'* - in these word-initial examples is followed by front vowels, while in the other two initial occurrences it is once followed by a front vowel (*\*δ'äŋ3-se* 'eine Art Gefäß aus Birkenrinde') and once by an unspecified vowel (*\*δ'Vkk3* 'stechen, stoßen') (cf. UEW, *ibid.*).

Denis Sinor, discussing PU locative and ablative suffixes states that, "local suffixes tend to change specific functions from one language to another or even within the same language, e.g. a locative may become an ablative or vice versa" (Sinor 1988:716). This statement is relevant for the functional diversity in the Nostratic daughter languages, especially in Altaic. He is confident that the PFU *\*-t ~ \*-d* ablative-locative is identical with a PA *\*-t ~ \*-d* ablative-locative that is reflected in Turkic and Mongolian as ablative-locative and in Tungusic as a dative (cf. Sinor 1976:126).

The identification of the phonological features of PU *\*δ* and *\*δ'* are still open to debate. Two suggestions seem to be feasible in the light of external comparison:

a) Janhunen's proposition that these segments are either related to the dental stop or to the liquids (cf. Janhunen 1982:24);

b) Honti - who does not approve of Janhunen's "Dentalspiranten" - suggests that these phonemes must have been lateral spirants of the type that survives in Ostyak (cf. Honti 1992: 211-212).

These ideas are not necessarily contradictory if they are considered in the context of Nostratic. As far as the above PU ablative suffix is concerned, the connection of PU \* $\delta$  with a dental stop is feasible (PN \* $d >$  PU \* $t$ -, - $\delta$ -), while the source of PU intervocalic  $\delta'$  could have been the PN lateral \* $\lambda$  (this may support Honti's lateral spirant) (cf. the table of Nostratic phoneme correspondences, Illič-Svityč 1971:149). The lateral origin of PU \*- $\delta'$ -, however, needs further etymological investigation and would deserve a separate discussion.

The phonological relationship between PU \*- $\delta a$ /- $\delta ä$  and PU \* $taka$  'rear' conforms to the established Nostratic sound correspondences according to which PN developed positional variants in PU: \* $t$  in initial position and \* $\delta$  in intervocalic position. These positional variants then phonologized in PU. The emergence of these allophones can thus be ascribed to the very process of grammaticalization: a clitic changes into a suffix and this modifies the morphonological situation as the original initial PN \* $d$ - finds itself in an internal position.

This is one of the points where we can again witness the heuristic capacity residing in Nostratic comparison. On the basis of Uralic internal comparison we would not be able to trace this grammaticalization process.

5. PD \*- $\ddot{t}tu$ /- $tt(V)$  'postpositive particle with locative-ablative meaning' (Illič-Svityč 1971: 212-213). On the Proto-Dravidian level it seems to be difficult, if possible at all, to reconstruct a locative/ablative suffix (cf. Andronov 1978:222). The ablative meaning is usually expressed by formants derivable historically either from postpositional collocations (frequently noun in a locative case) or from other, semantically reinterpreted case endings (cf. Andronov 1978:216). In many Dravidian languages, however, locative and ablative suffixes are derivable from a PD increment \*- $tt$ -, e.g. Parji ablative - $t$ -, or ablative - $t-ag$ , - $t-un$  (where - $t$ - is combined with dative suffixes), Naiki, Gondi ablative - $t-al$  (combined with an ablative suffix); locative suffixes, like Kolami - $t$ , Konda - $t(u)$ /- $d(u)$ /- $\#(u)$ , Bellari and Koraga -( $t$ ) $tu$ - $t$  (Andronov 1978:215-216, 221). Andronov considers this increment to be a relic of an ancient genitive-locative-ablative marker (cf. Andronov 1978:189, 202, 218).

Some of the examples cited in the Nostratic Dictionary have to be rejected (Kui, Kuvi ablative - $ti$ , Kuruh, Malto - $ti$ ) because in these languages these suffixes are not likely to be related to the PD increment \*- $tt$ - since they occur after the accusative suffix, whereas if they developed from an increment, they would precede other case suffixes. The circumstance that the PD increment \*- $tt$ - is closer to the root than other affixes may also point to its ancient emergence. The regular reflex of PN \* $d$ - in Dravidian is \*- $t(t)$ - but in the case

of the PD increment \*-*tt*- we are dealing with a bound morpheme that once used to be a free morpheme where the dental consonant used to be in initial position. This can account for the non-appearance of cerebralization.

6. PA \*-*da/-dä*, \*-*du/-dii* 'locative case formant' (Illič-Svityč 1971: 214). Vovin (1994:106) gives PA \*-*du*[*C*] > PMT -*duk*, PJ \*-*du*[*Ca*]. The Tungusic form with the final stop may be considered a reflex of the stop element in the original free morpheme (unless the Tungusic form proves to be a composite of two separate morphological elements). Illič-Svityč also lists Proto-Turkic \*-*ḍa/-ḍä* and \*-*ta/-tä* (after *r, l, n*) locative-ablative suffixes and Proto-Mongolian locative \*-*da/-de* surviving in adverbs (written Mongolian *en-de* 'here', *ten-de* 'there') and later developing a directional-dative meaning (which, by the way, reminds us of the functional developments Kartvelian!).

To sum up, the process of grammaticalization could have been as the following:

content word	> grammatical word	> clitic	> affix
PN * <i>daḲa</i>	> PN * <i>da</i> 'locative particle'	> PAA * <i>d</i> 'particle with locative meaning'	PIE *- <i>D/-eD</i> , PK *- <i>ad</i> PU *- <i>ḍa/-ḍä</i> , PD *- <i>ṭṭu/-tt(V)</i> , PA *- <i>da/-dä</i> , *- <i>du/-dii</i>

While the reflexes of the PN lexeme \**daḲa* survive only in AA, U, A and perhaps IE (Hittite) languages, we assume that the grammaticalized form of this lexeme was a more stable element - probably due to its affixed position (often preceding other suffixes and thus 'protected' from the tendency of word-final loss) - and is reflected much more extensively all over the Nostratic daughter languages (in fact in all six of what I like to call 'classical Nostratic', i.e. AA, IE, K, U, A and D).

I am indebted to Peter Michalove for several comments on an earlier version of this paper. I think the paper has improved as a result. He is not responsible, though, for the views expressed above.



## Abbreviations

- (P)A = (Proto-)Altaic  
 (P)AA = (Proto-)Afroasiatic  
 (P)D = (Proto-)Dravidian  
 PFU = Proto-Finno-Ugric  
 (P)IE = (Proto-)Indo-European  
 PJ = Proto-Japanese  
 (P)K = (Proto-)Kartvelian  
 PMT = Proto-Manchu-Tungus  
 (P)N = (Proto-)Nostratic  
 (P)U = (Proto-)Uralic  
 UEW = *Uralisches etymologisches Wörterbuch*  
 V = unspecified vowel

## References

- Andronov, Mixail S.  
 1978 *Sravnitel'naja grammatika dravidijskix jazykov*. Moscow: Nauka.
- Beekes, Robert S.P.  
 1995 *Comparative Indo-European Linguistics*. Amsterdam, Philadelphia: John Benjamins.
- Blažek, Václav.  
 1989 "Lexica Nostratica: Addenda et Corrigenda I", *Archív Orientální* 57/3: 201-210.
- Cohen, Gerald L.  
 1984. "Reflections on the Primitive Indo-European locative particles" *Zeitschrift für vergleichende Sprachwissenschaft* 97/1:13-21.
- Dolgopolsky, Aron (= Dolgopol'skij, Aron B.).  
 1971 "Nostratičeskie etimologii i proisxoždenije glagol'nyx formantov". *Etimologija* 1968:237-242. Moskva: Nauka.
- Dolgopolsky, Aron.  
 1992 "Nostratic etymologies and the origin of verbal formatives". In: *Nostratic, Dene-Caucasian, Austric and Amerind*. Ed. V.V. Shevoroshkin, 290-297. Bochum: Universitätsverlag Brockmeyer. [English translation of Dolgopolsky 1971].

Heine, Bernd - Ulrike Claudi - Friederike Hünemeyer.

1991 *Grammaticalization. A Conceptual Framework*. Chicago, London: The University of Chicago Press.

Honti, László.

1992 "Adalék a magyar *l* ~ finn *t* megfelelésének és alapnyelvi előzményének magyarázatához". [To the explanation of the Hungarian *l* ~ Finnish *t* correspondence and its proto-linguistic antecedents], In: *Rédei-Festschrift*. Eds. Pál Derék et al., 209-213. Wien, Budapest.

Hopper, Paul J. - Elizabeth Closs Traugott.

1993 *Grammaticalization*. Cambridge: Cambridge University Press.

Illič-Svityč, Vladislav M.

1967 "Materialy k sravnitel'nomu slovarju nostratičeskix jazykov" [Materials for a comparative dictionary of the Nostratic languages], *Etimologija* 1965: 321-372. Moscow: Nauka.  
1971 *Opyt sravnenija nostratičeskix jazykov*. Vol.1. Moskva: Nauka.

Itkonen, Erkki.

1969 "Zur Wertung der finnisch-ugrischen Lautforschung", *Ural-Altaische Jahrbücher* 41: 76-111.

Janhunen, Juha.

1982 "On the Structure of Proto-Uralic", *Finnisch-Ugrische Forschungen* 44: 11-22.

Kazanskij, N. N.

1989 "K rekonstrukcii kategoriji padeža v praindoevropejskom" [To the reconstruction of the category of case in PIE] in: *Atual'nyje voprosy sravnitel'nogo jazykoznanija*. Ed. A. V. Desnitskaja, 115-130. Leningrad: Nauka.

Klimov, Georgij A.

1964 *Etimologičeskij slovar' "kartvel'skix jazykov"*. Moskva: Institut jazykoznanija AN SSSR.

Lakó, György.

1968. *Proto-Finno-Ugric Sources of the Hungarian Phonetic Stock*. The Hague: Mouton.

Manaster Ramer, Alexis.

- 1993 "On Illič-Svityč's Nostratic Theory". *Studies in Language* 17/1:205-250.

Raun, Alo.

- 1988 "Proto-Uralic Comparative-historical Morphosyntax" in: *The Uralic Languages: Description, History and Foreign Influences*. Ed. Denis Sinor. Leiden, New York, København, Köln: E. J. Brill.

Rédei, Károly (ed.).

- 1986-91 *Uralisches etymologisches Wörterbuch. Vols. 1-3*. Budapest: Akadémiai Kiadó.

Shields, Kenneth

- 1987 "On the Indo-European Ablative" *Emerita* 55/1:63-69.

Sinor, Denis

- 1976 "The \*-t \*-d local suffix in Uralic and Altaic" in: *Hungaro-Turcica: Studies in Honour of Julius Németh*. Ed. Gy. Káldy-Nagy, 119-127. Budapest: Loránd Eötvös University.
- 1988 "The Problem of the Ural-Altaic Relationship" in: *The Uralic Languages: Description, History and Foreign Influences*. Ed. Denis Sinor, 706-741. Leiden, New York, København, Köln: E. J. Brill.

Szemerényi, Oswald.

- 1990 *Einführung in die vergleichende Sprachwissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Vovin, Alexander.

- 1994 "Long-distance relationships, reconstruction methodology, and the origins of Japanese". *Diachronica* 11/1: 95-114.

# Three Kisses<sup>0</sup>

Pramila Hemrajani  
*University of Michigan*

## 0. Introduction

In the debates over the nature of human nature that have so exercised scholars over the last century (and more), language has always been Exhibit A. But another behavior involving the use of the lips and the tongue has also figured in these epic disputations over nature (innateness, instinct) vs. nurture (culture, learning) and those concerning monogenesis (single origin) vs. polygenesis (multiple origins) of human behaviors.

It would take us too far afield to review the long history of this topic, but suffice to mention that it was no less an authority than Darwin (1872) who put forth the classic argument that the kiss (that is, the lip-kiss, for as we will see, there are also other kinds) is found only in some human cultures and hence cannot be innate. Incidentally, this view largely prevailed for a century, though it has been challenged more than once, most recently by another prominent scientist, Eibl-Eibesfeldt (1979:2, 1989:138-139 and *passim*), who holds that the kiss is a ritualized form of mouth-to-mouth feeding, a behavior "homologous in the great apes and man"--and hence presumably innate. And let us not forget that the kiss was the subject of one of the more elaborately argued hypotheses (advanced by the distinguished philologist Meissner 1934) positing the diffusion of a cultural trait out from a putative single place of origin (a hypothesis which is either ignored or rejected out of hand (e.g., Cooper 1983) now that "diffusionism" in general has fallen into complete, and not entirely undeserved, ill-repute in ethnology).

The difficulty of settling these fundamental issues (Is it universal? Is it innate? If not innate, was it "invented" once or more than once?) may itself be yet another parallel between kiss and language. In both cases, the questions are essentially the same, final answers are equally elusive, and the facts required to arrive at the answers are equally difficult to establish conclusively one way or the other. One classic example of this difficulty (among many) has been the inability of kiss researchers to decide just how old the kiss is in Japan. On the one hand, those who reject the universality of this behavior (e.g., Ellis

---

<sup>0</sup> This paper is based on research initiated in 1994 together with Alexis Manaster Ramer, to whom most of the strictly linguistic proposals reported here are ultimately due. Both of us owe a debt of gratitude to William Baxter, Belinda Bicknell, R. J. Campbell, Noam Chomsky, Peter Daniels, Gene Gragg, Jane Hill, Alexander Lubotsky, Jens Rasmussen, Gonzalo Rubio, Laurent Sagart, Brent Vine, and Alexander Vovin for discussion (variously) of certain points of Chinese, Japanese, Indo-European, Semitic, Sumerian, and Aztec philology, linguistic theory, and cognitive science. All opinions expressed here and any errors are, of course my sole responsibility.

1922:218) like to list Japan as a culture traditionally innocent of the kiss. But their evidence has been entirely anecdotal (primarily the observations of Hearn 1896:103, together with the fact that modern Japanese uses the English loan word *kisu*). On the other hand, those who believe in the innateness of the kiss (e.g., Eibl-Eibesfeldt 1989:138-139) naturally enough rejoin that kissing is mentioned in "old" Japanese texts discussed by Krauss and Satow (1965:368), hence proving that kissing existed in "ancient Japan". But the text in question refers to the tongue-kiss ("French kissing"), not the lip-kiss, and dates to 1695 (more than a century after the Portuguese arrived in Japan)--and the other citations in Krauss and Satow (1965) are undated. The evidence of Japanese erotic woodcuts is also not helpful: depictions of kissing are uncommon, and there is even the possibility that we are dealing with works created for export to Europe (Krauss 1911: 211). Likewise, the Longstreets' (1970:72) discussion of kissing in the pillow books from the Yoshiwara (the red-light district of Edo [Tokyo]) necessarily refers to the period no earlier than 1626, when the Yoshiwara first opened, hence after the first contact with the Portuguese. A possible indication of kissing before contact with Europe is the Longstreets' reference to (several) Japanese versions of the *Kāma Sūtra* (a work where kissing figures prominently) being commonly read in the Yoshiwara, yet we are not told anything of the date of these translations (not even whether they existed in the Yoshiwara in the 17<sup>th</sup> or only in the 19<sup>th</sup> century, for example). Moreover, even if this tantalizing lead were to pan out, it could merely point to an ultimately Indian (via Tibet and China), rather than a European, source for the custom of kissing in Japan (which, after all, has borrowed other things from India, e.g., Buddhism).

Although unmentioned in the many discussions of kissing (or absence thereof) in Japan, a more solid piece of evidence comes from the 10<sup>th</sup>-century work of Chinese medical lore, *Isinpō*, composed by Tanba Yasuyori, a Chinese physician living in Japan. The 28<sup>th</sup> section of this work is devoted to sex, and has been the subject of a fair amount of study, including translations into Japanese and English (e.g., Gulik 1961, Isihara 1967, Levy and Ishihara 1968). This section is a compilation of several even earlier Chinese texts (none of which have otherwise survived), and contains several references to what appears to be (based on the translations we have seen), once again, tongue kissing. This makes it absolutely clear that kissing (but perhaps only in this one form?) was not unknown in either China (as noted by Gulik) or in Japan (as apparently has not been noted) long before contact with Europeans. However, *Isinpō* contains several references to Buddhist texts (in its discussion of aphrodisiacs, of all things), and so once again we must deal with the possibility that all the references to kissing are also ultimately of Indian origin. It is even possible that kissing was not practiced (at all or at least by the vast majority of the population) but only read about (by a few literati). It does not seem unimaginable that its status at the time in Japan (and perhaps in China as well) was much the same as the status in our own culture of the celebrated positions of intercourse of the *Kāma Sūtra* and other non-Western sex classics: much

discussed but rarely emulated.

All in all, little can be concluded about the origins of the actual practice of kissing in Japan (or China) from the available arguments (or from the additional data of the *Isinpō*), and the same is true in just about every other such case that has been discussed in the literature. Invariably, the data are sparse and incompletely analyzed, and in any case involve points which, while interesting, are far from crucial (much as the mere demonstration that kissing was known in Japan before Commodore Perry's time is of little theoretical importance, considering the much earlier contact with the Portuguese (and the Dutch), not to mention the possibility of Chinese and Indian influence). Given such difficulties, which seem very similar to those encountered by linguists, perhaps the ban on speculation about the origin of language said to have been enacted in the last century by the Paris Linguistic Society should be extended to the ultimate origin of the kiss!

If so, then research on the kiss could concentrate on issues of more recent periods in history and prehistory (whose solution might over time eventually lead to deeper answers, of course), much as in comparative linguistics. Moreover, there are still more connections with language and linguistics involved in such work. For one thing, one of the most solid sources of evidence about the distribution of the kiss, and hence about its likely course of development across human cultures, is the careful study of the attestations and etymologies of the words for kissing in the different languages. For another, although this has almost never been mentioned in the literature on the kiss, it is of critical importance to determine what words or roots, if any, can be posited for the meaning 'kiss' in the various proto-languages linguists have been able to reconstruct. The ability to recover words in languages of which no records have survived allows us, of course, to penetrate much further into prehistory than is possible by merely considering the vocabularies of the oldest attested languages (such as Sumerian, Old Egyptian, Homeric Greek, Vedic Sanskrit, etc.), as most authors have done till now. The practice of linguistic analysis as a way to illuminate various aspects of culture denoted by the words analyzed is, of course, neither new nor entirely reliable. Automobiles, despite the mixed Graeco-Latin etymology of the word, were not invented in Antiquity somewhere in the middle of the Adriatic (the *auto*-part may be Ancient Greek, and the *-mobile* part Latin, but the combination was made up more recently and on dry land). But, with a modicum of caution, one can probably safely use etymological arguments to establish that the people from whose speech the Indo-European languages descend knew how to count at least to a hundred, were patrilineal and patrilocal, had domesticated animals and drank their milk, and so forth; that the Indo-Europeans learned much (if not all) about domestication from the Semites (Illich-Svitych 1964); that the Navajo are recent arrivals in the American Southwest (Sapir 1936); and so on. In the case of terms for kissing, very little work of comparable precision or scope has been done, and too often the conclusions that are drawn from the history of words have been too strong. For example, the Latin etymology of Welsh *pac* 'kiss' (from *osculum pacis* 'kiss of

peace') or the presence of an English loanword for 'kiss' in Japanese (namely, *kisu*) do not suffice to establish the borrowing of the custom itself. In fact, in the case of Japanese, we have attestations of several other terms in various dialects (Krauss and Satow 1965), and there must in addition have been *some* word for kissing in the language of the half-million or so Japanese who converted to Roman Catholicism between 1549 and 1612, (not to mention in the Sino-Japanese of Tanba Yasuyori a half-millennium earlier).<sup>1</sup> Finally, at least since the mid-19th century there has been another, less common, word for 'kiss' in Japanese: the Sino-Japanese term *seppun* corresponding to Mandarin Chinese *jiēwen*, which is said to be borrowed from the Japanese). Much work remains to be done on such problems for all languages and all culture areas, not just Japan. Below, I will discuss two particular cases in more detail, those of Proto-Semitic and Proto-Indo-European, and will try to argue that the method, if used sensibly, yields important results.

Finally, the tools of "dialect geography" (i.e., the methods used by linguists to decide which of two competing pronunciations or forms is older and which is newer just by studying their geographical distribution) can be adapted to the study of kissing, allowing us to determine which forms of kissing are older and which are newer in the same way. Thus, the discoveries made some 100 years ago by the founders of dialect geography can, as I will try show, cast an entirely new light on the prehistory of the kiss. (There is actually yet another connection between language and the kiss, namely, that the two leading figures who sought at the turn of the century to systematize what was known about the kiss, Nyrop and Siebs, and the author of the best articulated theory about its prehistory, Meissner, were all distinguished philologists, but that is a matter not for the students of the history of language or of kissing, but for the historians of science.)

Of course, while we will be focusing on the possible uses of linguistics, or methods borrowed from linguistics, in studying the (pre)history of the kiss, we should note that we are not walking a one-way street here. There are important ways in which research on the origin of the kiss can perhaps cast some light on the issues of innateness and monogenesis (or the opposite) which language, the traditional Exhibit A in such discussions, cannot (or at least has not). For example, in the case of the kiss, no one would seek to derive a claim of innateness from mere universality (as so many theoretical linguists seem to do, quite inappropriately, in the case of language). More importantly, the kiss researchers have at least begun to acknowledge, though usually without accepting (e.g., Eibl-Eibesfeldt 1979), the argument that innateness vs. learning and monogenesis vs. polygenesis are not entirely valid dichotomies, an issue which few linguists seem to be aware of. As we will see, this may be the most important thing which research on the kiss has to teach us: much more so than in the case of language, the study of the kiss leads one quite quickly to appreciate

---

<sup>1</sup> We have not so far been able to consult the original of the *Isinpō* in order to settle this question.

two relatively new concepts: (i) the possibility that various human behaviors have few but more than one origins ('oligogenesis'), and (ii) the need to recognize not only where a particular behavior came from in human (pre)history or what biological mechanisms allow it to be learned in each generation (the central questions, respectively, of comparative vs. theoretical linguistics, of ethnology vs. cognitive science, of anthropology vs. psychology) but also what biological mechanisms allow it to appear (once or more than once) in the first place (this being perhaps the "missing link" between these pairs of disciplines).

### 1. The Semitic Kiss

We begin with the Northwest Semitic (c.g. Hebrew) root *n-š-q* 'to kiss', in relation to the obviously related Akkadian synonym *n-š-q*. According to the usual rules for the correspondences of sounds among the Semitic languages, the Hebrew and the Akkadian forms would normally be related, implying a Proto-Semitic root *\*n-š-q*. This would mean that kissing could not have originated in Mesopotamia, as claimed by Meissner, since the Semites are not indigenous to that region.

However, a Proto-Semitic *\*š* has to correspond to Arabic *s*. This would then lead us to posit that this root is related to Arabic *n-s-q* (whose gloss is a nontrivial matter, as we will show directly). Such a connection has been widely assumed at least since 1812 (Gesenius 1859: 571). As for more recent work, in Köhler and Baumgartner (1948-53:640) we find, confusingly enough, **both** Arabic *n-s-q* and Arabic *n-š-q* (the latter meaning 'to smell, sniff, inhale; to snuff up the nostrils') cited as cognate with the Hebrew word for 'kiss', and moreover *n-s-q* incorrectly glossed '[to] fasten together'. Referring to this work, Cohen (1982), without any justification, further alters the gloss of the Arabic *n-s-q* from Köhler and Baumgartner's '[to] fasten together' to the subtly but crucially different 'to seal, fasten together', and on the basis of this argues that the meaning of Hebrew *n-š-q* originally referred to "sealing of the lips together". However, the meaning 'to be silent [i.e., to close one's mouth]' which Cohen posits to explain certain difficult Old Testament passages must be secondary, and certainly cannot be derived from the Arabic, where the root at issue is more properly glossed 'to string (pearls), to put in proper order, arrange nicely, range, array, order, marshal, dispose; to set up, line up'. It is precisely because of this meaning in the Arabic that some authors have recognized that any connection between Arabic *n-s-q* and kissing, in Meissner's (1934:918) words, "dem Sinne nach... nicht sonderlich gut paßt" ["does not fit the meaning particularly well"]--something of an understatement. Of course, semantic arguments can be tricky, but there is no way around the fact that Arabic *n-s-q* is related to Akkadian *n-s-q* (a *št*-stem) '[to] put in order, prepare' (Leslau 1987:403), hence cannot be cognate with Akkadian *n-š-q* 'kiss', and accordingly a relationship with Hebrew *n-š-q* is also impossible.



As first proposed by Lagarde (1887:24-25, cited in Barth 1893), the Akkadian and North West Semitic words for 'kiss' are therefore related, not to Arabic *n-s-q*, but, if at all, to Arabic *n-š-q*, which as noted means 'to smell, sniff, inhale; to snuff up the nostrils'. This connection, a priori equally difficult to credit, is rendered compelling by the fact that in many cultures the gesture equivalent to our kiss (i.e., the gesture used to express affection, especially between relatives or between lovers; to symbolize greeting, especially after a long absence; and the like) involves the use, not of the lips, but of the nose (e.g., Darwin 1872, Andree 1889, and the literature cited therein). This is the so-called custom of "rubbing noses", which in current Western cultural stereotypes is associated with the Eskimos, although 19<sup>th</sup> century scholars (including Darwin and Wallace, both of whom observed it at first hand) referred to the "Malay kiss", and even as recently as 1934 Meissner did not know of the existence of this behavior among the Eskimo. This form of kissing, more properly called the sniff- or nose-kiss, has been described by careful observers as involving a variety of different precise gestures (different in different cultures and according to some sources even between different Polynesian tribes) which only rarely have anything to do with rubbing but almost always do involve inhaling the odor of the other person.

In the geographical area of interest here, i.e., the Near East, the nose kiss is attested without doubt in Ancient Egypt, as well as in parts of Arabia today (e.g., Dostal 1983:63). It may indeed have been more widespread in the ancient Near East than the standard sources acknowledge. All too often words in the various ancient tongues which may have been ambiguous are assumed to refer to the lip-kiss, although, in reality, even in relatively recent texts (e.g., the Old Testament) we cannot be at all sure that all the kisses really involved the lips rather than the noses. For example, we read (Gen. 33:4) that Esau "fell on his [Jacob's] neck and kissed him" (for a similar turn of phrase in an Akkadian epic, see Speiser 1964:259). But how do we know what kinds of kiss this was? There is some temptation to compare these expressions to the usage found in Hawaiian legends where "persons swept by sudden passionate affection are described ... as "flying upon the neck" (*lele 'a'i*) of the beloved" (Pukui 1972) to perform a *honi* (a nose-kiss). We should not jump to conclusions, but the fact is that philologists analyzing references to kissing in the ancient texts usually do not even consider the possibility of the nose-kiss. Once that possibility is admitted, it seems likely that at least some passages in the extant texts from the ancient Near East will be reinterpreted as involving the nose-kiss.

The semantic connection between 'inhale' and '(lip-)kiss' can now be easily explained. There must have been a missing link, the hypothetical meaning \* 'nose-kiss'. This sense is unattested for this root in those parts of Arabia where the nose-kiss is still practiced today (Dostal 1983, describing the usage in the north of the United Arab Emirates, gives the term *xašm*) or indeed in any Semitic language (unless of course some of the "kisses" referred to in ancient Near Eastern texts really were nose-kisses, as I suggested). Yet we must posit it for the prehistoric form of (what became) Akkadian. A semantic shift to

'lip-kiss' from 'nose-kiss' would be perfectly natural. It clearly occurred in Ancient Egyptian (and if I am right elsewhere in the Near East) and in the modern Eskimo languages. Moreover, there are ethnographic examples of a parallel shift from the custom of nose-kissing to that of lip-kissing, but none known of the reverse. Finally, the derivation of 'nose-kiss' from 'inhale' seems trivial. Hence, it makes sense to assume that in Proto-Semitic this root had the sense 'inhale', still preserved in Arabic, and that the sense of 'lip-kiss' known to us from Akkadian and Northwest Semitic was derived via **two** independent semantic shifts.

However, now we have a problem with the sibilants. We know from many examples that words where both Akkadian and Arabic have *š* correspond to Hebrew words with *ś*. To avoid confusion, we can summarize the relevant facts as follows:

Proto-Semitic	* <i>ś</i>	* <i>š</i>
Akkadian	<i>š</i>	<i>š</i>
Hebrew	<i>ś</i>	<i>š</i>
Arabic	<i>š</i>	s

In short, if the Akkadian and Arabic forms are related, then the Hebrew one cannot be. Barth (1893) suggested that the irregular *š* in the Hebrew could be due to the presence of a velar sound (in this case *q*) in the same root. If there were a general law whereby Hebrew has *š* in place of *ś* next to a velar, that would, of course, solve the problem, and the irregularity would be only apparent. However, there are attested examples of Hebrew *ś* in the same word with a velar, e.g., *kaśa* 'became fat', *kebeś* 'sheep', and so on. All this did not unduly trouble Barth, who clearly regarded the putative change of *ś* to *š* next to a velar as a tendency rather than a law, and in general did not hold to the doctrine of exceptionless sound laws, which at the time had just recently been proposed and was being vigorously debated.

However, the idea of resistance (or exceptions) to sound laws is a very nebulous and controversial one, and recently more and more apparent examples seem to crumble on closer inspection. Moreover, it is not so much a question of whether such exceptions occur, as of linguists' having to base their conclusions about linguistic prehistory (and that includes etymology) on data which exclude such exceptions. While exceptions to sound laws may actually occur, especially in the case of baby-talk, onomatopoeia, and ideophones, this simply means that such cases are not amenable to analysis by the conventional tools of comparative linguistics. Words whose prehistory involves deviations from sound laws remain unexplained precisely to the extent that they involve such irregularities, for it is the regularity of sound laws which is the principal protection linguists have against incorrect etymologies (Manaster Ramer, to

appear b).

There is, however, another possible explanation for similar (or identical) sounding words with similar (identical) meanings which is amenable to analysis by comparative linguistics: borrowing. The Semitic facts can most naturally be explained as a borrowing of the Akkadian word into Northwest Semitic (something of which there are also many other examples). Such a scenario (semantic shift in Akkadian, then borrowing from Akkadian into Northwest Semitic) would explain **both** the distribution of the meanings ('kiss' vs. 'inhale') **and** the questionable sound correspondences. As for sounds, if the Hebrew word (with other Northwest Semitic cognates) is **not** derived directly from a Proto-Semitic etymon (but borrowed from Akkadian), then what looked like a violation of the sound laws is instead the natural consequence of the borrowing process. Hebrew has *š* instead of *s* because it got the word from Akkadian, which is supposed, quite lawfully, to have *š*. As far as the meaning is concerned, the borrowing hypothesis makes it unnecessary to assume what would otherwise seem inescapable, namely, that the same sequence of two separate semantic developments occurred independently in Akkadian and Northwest Semitic. Moreover, both the direction and geographical location of the semantic change posited as well as the direction of the hypothesized borrowing all make sense given the little we do know about the cultural patterns concerning lip-kissing and sniff-kissing in the Near East and in general. It makes good sense to think that the speakers of Proto-Semitic did not know the custom of joining the lips, that it was the Semitic tribes who ended up in Mesopotamia who first changed over from nose- to lip-kissing (a custom which is attested among non-Semitic Mesopotamians, namely, Sumerians), and that this custom was then learned from them by other Semites.

While we have just barely scratched the surface, we thus see that linguistic evidence, analyzed in more detail than hitherto (although surely more can be done by experts in the field), yields some rather useful results--results which in this one case at least tend to support the hypothesis of Meissner (1934) about a Mesopotamian origins of the lip-kiss and its subsequent diffusion to other culture areas (even if Meissner himself did not draw the same conclusions about the developments of the medial consonant of *n-š-q*). It is of particular interest that, while this root (based on its occurrence in Akkadian and Arabic) can be reconstructed for Proto-Semitic, there is no basis for positing that the meaning 'lip-kiss', found in Akkadian and Northwest Semitic, dates to Proto-Semitic. Indeed, there is absolutely no reason to believe that Proto-Semitic had **any** word or root for the lip-kiss, inasmuch as this root is (or, rather, was) the only conceivable candidate for such a role. The various other roots found for this meaning in the various languages clearly involve local developments. This is important because the absence of a word with such a meaning in Proto-Semitic is logically a necessary consequence of Meissner's diffusionist theory. Although one can never prove a negative, our inability to reconstruct such a Proto-Semitic word could be part of a broader case for his hypothesis.

## 2. The Indo-European Kisses

While this topic is only rarely discussed in the literature on the kiss, it is also of interest to determine whether we can reconstruct a Proto-Indo-European word for 'kiss'. For one thing, if the answer were in the affirmative, then this would in effect suffice to refute the claims that the pre-Homeric Greeks (Meissner 1934:930, with references to earlier work) and the Vedic Indians (Hopkins 1907, Meissner *op. cit.*), did not practice kissing. The absence of an inherited Indo-European word for kissing in most Indo-European languages (including Sanskrit; but see below) would then simply mean that such words were replaced in the course of time by borrowings or neologisms, with no prejudice to the antiquity of the custom of kissing itself.

Much as in the case of Semitic (but more so, because we are now dealing with a much more diverse language family), there are many different and unrelated words or roots for the (lip-)kiss in the various Indo-European languages. But, as in the case of Semitic, there is only one which could possibly be a candidate for Proto-Indo-European status. The Indo-Europeanist literature often posits such an etymon, written as *\*ku-*, *kus-* by Pokorny (1959:626). This notation appears to be intended to lump together a stem ending in *\*/s/*, which is attested in three branches of Indo-European (specifically Greek *κυνέω* with the aorist *έκυσ(σ)α*; Hittite *kuwašš-* (*/kwas/*); and Germanic *\*kus-*, as in English *kiss*, German *Kuß*, etc.) and a purely Germanic *\*kuk-* (as in Gothic *kukjan*), and to facilitate a somewhat arbitrary connection he hints at with a sound-imitative root he writes as *\*bu-* (p. 103), although this does not really account for most of the forms. Pokorny further complicates matters by citing as "ähnlich" ["similar"] some Sanskrit forms with meanings like 'to suck' or 'to make noses while eating' which begin with *c /č/*, a sound which is impossible to derive from PIE *\*/k/* in this position (the same applies to the relationship sometimes posited by some Indo-Europeanists with (late) Sanskrit *cumb-* 'to kiss').

Even if we omit these questionable connections, and stick to just *\*kus-*, as most recent investigators seem to, the problems do not end. In particular, the Greek and the Germanic words cannot be related under the known sounds laws which would require that either (a) if Germanic has */k/*, then Greek should have */g/* (implying a PIE shape *\*gus*) or (b) if Greek has */k/* (implying PIE *\*kus*), then Germanic should have */h/*. For it was one of the earliest discoveries of comparative linguistics ("Grimm's Law") that Germanic languages have invariably replaced Indo-European *\*/k/* sounds with */h/*. It should be noted that Hittite has */k/* equally for PIE *\*/k/* and *\*/g/*, and hence cannot help us here. The standard attempt at an explanation of the discrepancy between the Germanic and the Greek has been that the Germanic word resisted the sound change because it is a "Schallwort", that is, a sound-imitative word (onomatopoeia). However, this hypothesis (which too often, e.g., by Pokorny himself, is presented as though it were fact) has never been properly defended.

The most obvious difficulty is that, once the regularity of sound laws

in no longer assumed, it is just as easy to posit that the root was PIE *\*gus*, with an irregular devoicing in Greek, as it is to reconstruct *\*kus*, with an irregular failure of Grimm's Law.

Next, it would have to be shown, not merely asserted, that these words were in fact onomatopoeic. Real onomatopoeia, although somewhat language-specific, tend to be rather similar across unrelated or distantly related languages. If one wanted to argue that Sanskrit *cumb-* 'to kiss' is onomatopoeic, one might cite the strikingly similar Polish baby-talk terms: the ideophone *cium*, often reduplicated (*cium-cium*, cf. English *kiss-kiss*), and the related verb *ciumać*. Moreover, the real onomatopoeia for kissing seem to typically contain a labial sound, as in the examples from Sanskrit and Polish, in English *smack*, German *Schmatz*, and the like, and in the various scattered Indo-European forms which led Pokorny to posit *\*bu*, such as English *buss*, German *Buss*, and the like), etc. But nothing like this appears to be the case for the sequence /kus/ or even more generally /kVs/. The closest non-Indo-European analogue to a word of such a form may be Sumerian KI-(A) SU-UB (which we write this way to emphasize the fact that Sumerian phonological values are not fully established). But, however this was pronounced, it is a transparent phrase meaning 'to kiss the ground', more precisely 'to press the ground (with the mouth/nose)', in contrast to the expression for kissing in general, NE SU-UB (Cooper 1983:377), and not an onomatopoeia for kissing at all.

Third, while it is common for ideophonic or baby-talk forms (whether strictly onomatopoeic or not) to exhibit phonological anomalies, it is less clear that words derived from them do. It is instructive to compare Latin/French *caca* /kaka/ 'poop', which is an exception to the various relevant sound changes from Latin to French, with the derived verb (Latin *cacare*), which no longer resists these same sound changes and so becomes French *chier*. The Germanic root *\*kus-* and its putative cognates in other Indo-European languages are all attested in derived verb forms, which should presumably have been subject to sound changes such as Grimm's Law. It would not be surprising, then, if an ideophone like English *kiss-kiss* defied Grimm's Law, but the verb *to kiss* (and the noun *kiss*) should have undergone the change to /h/.

Finally, as noted above, the refusal to insist on the exceptionlessness of sound laws means the abandonment of the best criterion linguists have for distinguishing real connections from spurious ones. Words whose prehistory involves deviations from sound laws remain unexplained precisely to the extent that they involve such irregularities, for it is the regularity of sound laws which is the principal protection we have against incorrect etymologies. In the case before us, if Germanic *kus-* derives from a PIE *\*kus-*, and all we can say about the /k-/ is that it is irregular (and moreover cannot even justify the assumption that this is due to the allegedly onomatopoeic character of the word), then we really have very little basis, if any, for positing that the Germanic word really derives from the PIE one at all. And as a matter of fact, Meissner (1934:930) quotes personal communication from the well-known Indo-Europeanist Wilhelm Schulze as arguing that the Germanic and the Greek words are only related by

"Elementarverwandschaft" (literally "elementary kinship"), i.e., that there was no connection via a common ancestral root in a reconstructable protolanguage but merely the kind of "kinship" which exists between like-sounding onomatopoeia, ideophones, or baby-talk words in unrelated or distantly related languages (which is either no relationship at all, as I am told most linguists now assume, or perhaps a genuine relationship, but one involving a mechanism of cultural transmission separate from the transmission of (the rest of) language and hence subject to different laws).

However, as we just saw, there is no compelling reason to invoke onomatopoeia or the like in the case of the words under discussion, and so "Elementarverwandschaft" is not an attractive explanation. On the other hand, as we saw in the Semitic case, there is another explanation for similar (or identical) sounding words with similar (identical) meanings besides relationship and "Elementarverwandschaft", namely, borrowing. Given what has been said, it seems to make good sense to assume that either the Greek or (more likely) the Germanic word is not directly inherited from Proto-Indo-European but instead represents a borrowing. If the PIE root began with \*/k/, then Germanic could have borrowed this word from a language that did not undergo Grimm's Law. If the PIE root began with \*/g/, then Greek could have borrowed it from a language that did. Such a borrowing hypothesis would have the advantage of being consistent with the thesis that sound laws are exceptionless and with the fact that words for kissing are often borrowed, and that there is no indication whatever that \*kus- was onomatopoeic.

But now we would have only two witnesses to the erstwhile existence of this PIE root (since one of the three would be a borrowing). This is not very encouraging: Meillet argued that one should normally have three witnesses for any linguistic reconstruct (although I am told he did not always stick to this rule of thumb). Still, if there were no further problems, one would probably cheerfully accept the reconstruction of either \*gus or \*kus 'kiss' for Proto-Indo-European, contradicting Meissner's theories. However, in addition to all these difficulties with the voicing of the initial consonant of the putative PIE root (\*/k/ or \*/g/), there are still other problems which make it somewhat doubtful that there was a PIE root of this shape at all, even one reflected only in Hittite and in either Greek or Germanic (but not both).

The first of these problems involves another aspect of the initial consonantism. Although everybody writes it as \*/k/ (the PIE plain velar), it could just as easily have been \*/k̟/ (the PIE palatovelar), since this root is not attested (but see below) in any language which distinguishes velars and palatovelars. There is also a problem with the vocalism. Pokorny's \*kus- is a so-called zero-grade form of a root which could have any one of a number of full-grade shapes, the exact number differing according to which version of Indo-European theory one believes. These certainly include \*kews, \*kwes, \*kwos, and \*kows as well as (once we take the different possibilities for the initial stop) \*k̑ews, \*k̑wes, \*k̑wos, \*k̑ows, \*gews, \*gwes, \*gws, \*gows, \*g̑ews, \*g̑wes, \*g̑ws, and \*g̑ows, for a total of 16 distinct possibilities. As the number of

possibilities grows, so does, of course, the likelihood that the attested words come from distinct etyma.

Moreover, it is far from clear how exactly the Hittite form /kwas/- was supposed to be connected to any of these. This may have lain behind Meissner's refusal to admit that the Hittite word was related to the Greek or the Germanic at all, in fact. A half-century on, we now have the widely accepted reconstruction of this root by Eichner (1988) as *\*kuas* (i.e. *\*kwas/*), based precisely on the Hittite evidence (it could, of course, equally well be *\*kwas*, or, if we relate the Hittite to the Germanic rather than the Greek, *\*gwas* or *ǵwas*). This cuts down on the number of possibilities quite significantly, and makes sense of the Hittite form. Eichner's proposal, along with the whole idea of an *\*/a/* vowel in PIE is controversial, especially inasmuch as one of the leading groups of Indo-Europeanists, the so-called Leiden school, has made it their business to try to explain all the alleged cases of *\*/a/* in PIE as reflexes of the second laryngeal or of *\*/o/* (see, e.g., Schrijver 1991:4). However, if Eichner's reconstruction is accepted, then it becomes relevant to consider that many authors have long argued that PIE *\*/a/*, while perfectly real, was typically found in loanwords (and in certain highly circumscribed semantic domains such as in words denoting physical defects). Thus, even if the reconstruction were correct, it might simply mean that we are more than likely dealing with a loanword. To be sure, this borrowing would have occurred at a much earlier time than Meissner would have surmised (i.e., not into Greek or Sanskrit but into Proto-Indo-European). This would force some changes in the diffusionist theory of the kiss, but the basic idea of diffusion (with Indo-Europeans being on the receiving end) would be upheld.

On the other hand, like some of Eichner's other examples, the *\*/a/* in PIE *\*kwas/* is extremely tenuous, precisely because the evidence for the *\*/a/* is exclusively Hittite. This is thus an especially glaring example of a reconstruction based on not on three witnesses (as per Meillet) and not even on two, but on one. Logically, the *\*/a/* should perhaps not be able to be posited for PIE in this situation at all. This is, I am told, why comparative linguistics is **comparative**: since anything found in only one language may have evolved in that one language, linguists can only reconstruct proto-languages by **comparing** two (or preferably three or more) languages with each other. As things stand, the */a/* in Hittite /kwas/- cannot be compared to anything, and hence it could represent an innovation proper only to Hittite (or perhaps to Anatolian, the branch of Indo-European to which Hittite belonged), and so could the whole pattern of */a/-grade* verbs forms like it, such as */hwap/-* 'to injure', etc. (a pattern which does not seem to have any parallels elsewhere in Indo-European). Given how much the Indo-Europeanists' ideas of the rules of Hittite phonology have been evolving recently (e.g., Melchert 1994), it seems premature to assume that there may not at some point have been a hitherto-unformulated minor rule which would derive the */a/* in such words from something else. (Manaster Ramer conjectures *\*/e/*.)

If this were correct, and if, furthermore, we focused on the possibility

that the initial stop was \*/k/ rather than \*/k/, then a possibility to consider would be that we are dealing (for Hittite and Greek, anyway) with derivations from a much better-established PIE root: \**kues-*, \**kus*, 'keuchen, schnaufen, seufzen' ['to wheeze, breathe heavily, sigh'], as glossed by Pokorny (1959:631), or perhaps (as in the Sanskrit and Avestan derivatives) simply 'to breathe'. The semantic connection would be made on the basis of the ethnographic evidence that it has often been believed, as among the Maori, that "When two people greeted each other by pressing noses in the hongi ['nose-kiss'], they were intermingling their *hau* ['human breath, vital essence']" (Orbel 1985:75). Similar beliefs are attested in cultures which practice the lip-kiss, too, so even if the derivation suggested here for the Indo-European words for kissing were correct, this would not necessarily prove that the Indo-Europeans originally kissed by touching noses rather than lips, but in any case it would mean that we could no longer posit a PIE etymon with the original sense 'kiss'.

To be sure, there is yet another possible IE connection, which points to an entirely different derivation, but which, once again, would make the 'kiss' sense secondary. The Sanskrit grammatical literature lists a root *kus-* (also given as *kus*) 'to embrace'. Since this is not otherwise attested, and since the indigenous grammatical literature abounds in made-up roots designed to "explain" various words of obscure etymology, this form is usually ignored in the Indo-Europeanist literature, though not in the literature on the kiss (e.g., Lombroso 1893). But if this root was real, and not a grammarian's invention, then we might want to relate it to the Greek (and Hittite) forms (the highly unusual /s/ after /u/ in the Sanskrit would be explained if the PIE root were \*/kwes/, \*/kwas/, or \*/kwos/, and the /s/ in the zero-grade were analogical to the full grade). The Sanskrit data would then tell us that the PIE root began with \*/k/, not \*/k/ or \*/g/ or \*/g/, and that its original sense presumably had nothing to do with the nose-kiss. But in that case (as in French *embrasser*), the 'embrace' sense is more likely to be primary than is the 'kiss' sense, and so once again we would have to give up the idea of a PIE root with the latter meaning.

Of course, if either scenario were true (that is, if the 'kiss' sense is derived from some other primary meaning within Indo-European, whether 'breathe' or 'embrace'), it would then become crucial to determine whether the semantic evolution to 'lip-kiss' took place independently in Greek and in Hittite, and whether it was influenced in any way by non-Indo-European languages. The fact that the only two Indo-European languages to exhibit this development (since the Germanic would still have to be a borrowing) were spoken in relatively close proximity to each other (the Greeks and the Hittites are thought to have run into each other during the Trojan War, the Hittites being apparently the original referent of the term 'Amazon') and to the culture area claimed by Meissner to be the original homeland of the lip-kiss might then be quite significant. This is speculation, of course, although perhaps no more so than some of the standard proposals advanced hitherto regarding PIE \**kus* or \**kwas* 'to kiss'. For now, the most that we can say is that the case for a PIE root, of



native origin and denoting (lip-)kissing, is not a strong one, although the possibility cannot be excluded entirely, and that we can see a number of quite specific issues that need to be confronted before the final answers can be given (by the experts in Indo-European, of course, but not without due attention to the ethnographic and linguistic parallels from other language families and culture areas). Thus, this crucial case for the different theories of the origin of the lip-kiss, while it looks promising for the diffusionist side, is far from established one way or the other.

### 3. The "Dialect Geography" of the Kiss

We have so far focused on the linguistic evidence bearing on the origins of the words denoting the kiss, but as noted at the outset, there is an even more important contribution linguistics can make to this subject. The distribution of ethnographic differences involving different kissing behaviors yields interesting results when analyzed using the methods developed by the students of dialect geography.

While dialect geographers study differences in pronunciation and other linguistic features, we are concerned with differences between cultures having to do with precisely **how** they kiss. Much of the debate about the origin of the kiss has been confused by lack of clarity about which of the many different behaviors which can be called 'kissing' are at issue. For example, we will have to distinguish not only the lip kiss from the nose kiss (the one distinction that is usually made in the current literature), but also from various kinds of "kissing" involving the tongue and the teeth (both of which appear to be better attested in traditional East Asia than either the lip or the nose kiss, for example). Next, we will need to be careful to document just what significance any given kind of kissing has in a particular culture, e.g., whether it is a gesture of greeting/farewell, a sign of affection, a way to cajole or seduce, a legal step in making a contract, a religious gesture or even sacrament, a sex act, etc. For example, Eibl-Eibesfeldt's celebrated footage of mouth-to-mouth feeding and related gestures, recorded in various tribal societies outside the areas I described above as being characterized by having the lip-kiss, may--or may not--have a relation, despite his claims, to the lip-kiss proper. Likewise, it is necessary to consider the salience that kissing has in a given culture, i.e., how frequent it is, how public, how much it gets commented on by the natives of the culture, and so on (e.g., oversimplifying enormously, it seems that Herodotus testifies to the Persians--and Plutarch to the Romans--being more into kissing than the Greeks). Finally, we will need far better ethnographic (and linguistic) data about such distributions than the (largely anecdotal and amateur) reports we rely on at present.

Above all, there has not been nearly enough of an effort to map the (geographical or historical) distribution of these different behaviors across all the different cultures. Africa, pre-Columbian America south of the Eskimo belt, and Australia and New Guinea appear to be particularly poorly studied in this regard,

although no culture has been studied systematically. As a result, any attempt to identify broad patterns is fraught with difficulties. Yet, despite all these limitations, there are some things that seem tolerably clear, and which, as I promised, seem to become significant in light of the methods of dialect geography.

First, we know of many instances where the lip-kiss has clearly superseded the sniff-kiss in a particular culture, but apparently none of the reverse. A linguistic generalization is also possible: words for the lip-kiss often derive from words denoting sniffing, inhaling, or the like, but apparently never the reverse (if we except the scholarly coinage of terms like 'sniff-kiss', 'nose-kiss', and the like, as a way for Westerners to refer to Eskimo, Polynesian, and so on customs of this sort).

Second, the areas where lip-kissing is clearly a traditionally well-known form of behavior constitute a compact block encompassing all of Europe, the Near East (together with North Africa and Central Asia), South Asia, and Tibet. The areas where the nose-kissing is well-attested form a more complex pattern, including attestations in Ancient Egypt, among the Bedouin, in various places in Africa, all along the circumpolar belt (from Lapland through Greenland), and in Southeast Asia and the Pacific (i.e., from the hills of Assam to New Zealand to Hawaii), but apparently not among the Ainu or the Japanese. In other words, the typical lip-kissing areas seem to be contiguous and central, whereas the nose-kissing areas are scattered and peripheral, which, as we will see, may be significant.

Third, to the extent that it is possible, with our poor sources of information, to discuss such things at all, we may venture a generalization that, out of those areas where lip-kissing is indigenous (i.e., found as far back in time as our knowledge now reaches), some are ones where (lip-)kissing has been culturally salient (i.e., where it occurs frequently, in a wide variety of situations and forms, figures prominently in art, poetry, and the like, etc.)--and some are not. Although this generalization is based on extremely fragmentary and often unreliable information, there may be some basis for concluding that kissing has, in historical times, been more culturally salient in the Middle East (together with those parts of Europe abutting on the Mediterranean) than elsewhere in the lip-kissing parts of the world.

If all this is true, then this would immediately call to mind the discovery of dialect geographers that (a) older linguistic patterns tend to survive in scattered, often peripheral areas, whereas innovations tend to have a single center from which they radiate outwards and hence to occupy a contiguous area around such a focus, and that (b) there is typically more salience, complexity and variation involving any particular phenomenon in those places where it is old than where it is new (inasmuch as complexity and variation take time to develop). Tentatively, then, the generalizations we just made about the lip-kiss would lend some support to the diffusionist theories of authors such as Meissner. The sniff-kiss would be a form of behavior which must once have occupied much or perhaps all of the territory where it has long been superseded

by the lip-kiss. The custom of lip-kissing, wherever (or, as we will show below, almost wherever) we find it long-established and well-attested in historical times, would be a recent and (almost) unique innovation in human behavior, originating somewhere in the Near East, whence it spread to Europe (and in modern times throughout its universal cultural empire) and South Asia and Tibet. The case of East Asia is still rather unclear, even though tongue-kissing in China seems to antedate contact with modern Europeans by over a thousand years (see above), because kissing seems far less salient there than in other areas (e.g., d'Enjoy 1897), and because one might plausibly suppose that it came to China either from South Asia via Tibet (the earliest attestation cited by Gulik 1961, the one preserved in *Isinpō*, comes half a millennium after the arrival of Buddhism in China)--and/or from the Near East via Central Asia, another possibility that deserves closer investigation, given the massive influence on China from that quarter.

And so we come to yet another connection with language. There seems to be a pendulum which swings through all the disciplines of the humanities and social sciences. In the last century and the first part of this one, unforgivable excesses were committed by "diffusionists" of various kinds in every such field (and we may include here many of the early advocates of linguistic monogenesis or other similar linguistic theories). The reaction in all these fields was a predictable retrenchment, a shying away from attempts to reduce the bewildering diversity of human culture and behavior in historical times to any simple schemata. But now the pendulum is on the move again, with the work of linguists like Illich-Svitych and Greenberg the best-known examples, but not without parallels in other fields. There now seems to be a growing body of evidence that many human behaviors, including writing, counting (or at least counting above two or three or so), the use of rhyme in poetry (Greenberg 1960), the use of clicks as consonants in language (Manaster Ramer 1989), and of course (if we accept the recent work on "remote relations") language itself, had a single origin.

Or perhaps, as we are about to see, a very few distinct origins.

#### 4. Conclusion and Prologue: Oligogenesis?

The reason why I referred above to the possibly single origin of "lip-kissing, wherever we find it long-established and well-attested in historical times" is that it is one thing to demonstrate (as may be doable) such a common origin for many, even most **attested** instances of some behavior. But it would quite another (and this seems logically impossible) to show that there was not even a single other time in the whole (pre)history of mankind that the same behavior was "invented" independently. Lip-kissing, once it arose in Mesopotamia, may have spread through much of the world, but it is logically possible that there were other, independent origins, none of which had the same kind of "success" as the Mesopotamian developments, and, because they were

not as "successful", left fewer traces for us to discover.

This problem is not restricted to work on the kiss. All the results in the different fields alluded to above which are presented as demonstrating a single origin and diffusion from there actually seem to involve a paradox. On the one hand, in terms of the theoretical questions about the nature of human behavior the crucial distinction is seen precisely as being between one origin (monogenesis) and more than one (polygenesis). On the other hand, in terms of empirically determining how many origins of any phenomenon there have been, it is difficult (or perhaps even impossible) to distinguish between one and other small numbers such as two or three. In the case of rhyme, for example, Greenberg persuasively traces almost all instances of it to a Middle Eastern source, but he is not as confident that its occurrences in East Asia are historically related. It may be clear that rhyme did not originate independently in the hundreds or thousands of poetic traditions where we now find it, but it is exceedingly difficult to determine if it originated precisely once, as opposed to twice or thrice. Writing, too, in most of its known occurrences can be traced to a single Middle Eastern source, but it must have sprung up independently once or twice more (or so) in Mesoamerica and elsewhere. It is also easy to argue that the lip-kiss did not originate independently in each of the cultures that practice it, or even in most of them, but it is hard to see how we could show that it arose only once (as opposed twice or thrice) in human prehistory. It may well be precisely this discrepancy between what we would like to know and what we can know that accounts for the frustrating state of the debates about the origin and nature of human behaviors. It would then be nice if the paradox could be resolved. Perhaps it can.

Perhaps the more important distinction is in fact that between (very) few origins and (very) many, not that between one and more than one, after all. The thing to contrast polygenesis with is thus not monogenesis, but oligogenesis. Note in particular that, if a given behavior is innate, this does **not** mean that it has to be universal (members of the same species do differ in their genetic endowments), and even if it is universal, this does not mean that it has to have a single origin. For one thing, the same mutation can occur, and spread, more than once in the evolutionary history of a species. Researchers into the origins of our own species are still vigorously debating whether *Homo sapiens* emerged in just one (African) locale or in more than one. For another, the same behavior may arise via more than one distinct mutations. But it is very unlikely to have many distinct origins, because the chances of the same (or similar) biological accident occurring  $N$  times decline rapidly as  $N$  grows. Hence, the theoretically significant distinction is, after all, not that between mono- and polygenesis but that between oligo- and polygenesis (with monogenesis being just a special case of oligogenesis).

Moreover, the same general conclusion applies in the case of phenomena which are **not** innate. Here, the emphasis on oligogenesis correlates with a recently proposed idea for redrawing the line between innate and learned behavior (Manaster Ramer 1989, to appear a). As Chomsky has argued over the

last several decades, it is a conceptual necessity to view even learned behavior as involving a biological component, an "acquisition device" (presumably the same thing as Darwin's (1871) "instinctive tendency to acquire an art"). However, Darwin only posited this in the case of language, and drew a sharp distinction between language and other learned behaviors. Chomsky's early work, emphasizing as it did that language was not really "learned" but rather "acquired", followed suit, but more recently Chomsky (1975:24-25) took the first steps towards suggesting that other learned behavior also involves this kind of "acquisition". To be sure, he still assumes that, unlike language, there are other "domains" which "fall outside of [a human being's] cognitive capacity" and that in such cases "we will not expect [a] person to be able to find or construct a rich and insightful way to deal with the problem, to develop a relevant cognitive structure in the intuitive, unconscious manner characteristic of language learning and other domains in which humans excel", and that accordingly "Humans might" only "be able to construct a conscious scientific theory dealing with problems in the domain[s] in question, but that is a different matter". But Chomsky then adds the crucial qualification, "or better, a partially different matter, since even here there are crucial constraints" and offers some remarks on the "human 'science-forming' capacity". According to Manaster Ramer, however, there is no difference between the way we must study human language, science-forming, or any other capacities, although empirically it may turn out that there are huge differences between the different capacities which will emerge as these become better understood. However, the dichotomy between "acquired" behaviors like language and "cultivated" ones like, for example, science is rejected.

In addition, Manaster Ramer argues that a theory positing an "acquisition device" (an "instinctive tendency to acquire") can only explain the ontogeny of a given behavior (its development in an individual who is surrounded by others who already have it). There must also be another "instinctive tendency" (or "device"), one which is capable of producing a given behavior without role models and hence of allowing it to emerge phylogenetically. Although it is conceivable that the two mechanisms are ultimately the same, this is an open question (Manaster Ramer 1989 argued that, in the case of language, they were not, but his conclusions seem too strong). The reason this is related to the mono- vs. oligogenesis issue is simple. Once we accept Chomsky's as well as Manaster Ramer's ideas on the innate mechanisms involved in learned behavior, the notion of monogenesis loses all meaning, at least in relation to such behavior. This is because behavior which came into existence once could have done so more than once if the innate mechanisms which made it possible were in place. On the other hand, the notion of oligogenesis is still significant, because an event such as the appearance of lip-kissing--or of language--must have occurred under some very special, favorable circumstances. This is crucial, for the various mechanisms operate only under the right conditions ("with the right inputs", in the terminology of computer scientists and linguists). Since all behavior is seen as

having an innate component, the real question is how (i.e. under what conditions)--and in particular how easily--any given learned behavior emerges among human beings. The reason the important distinction is that of oligo- vs. polygenesis is that the former may be a sign that the mechanism is highly sensitive to external conditions ("very picky", one might say) and hence likely to have a rich innate structure. Chomsky's Darwinian dichotomy between "intuitive, unconscious" behavior, such as one's native language and "construct"ing a "conscious ... theory" of "a domain D that lies outside of [human] cognitive capacity" is thus naturally replaced with a view which allows different degrees (and kinds) of accessibility of a given behavior--whether we are dealing with individual acquisition or with the rise of the behavior within human culture.

All this finally means that the study of the (pre)historical origins of any given form of behavior has a direct bearing on the understanding of the biological mechanisms. This may fly in the face of the opinion of many in fields such as linguistics, where the study of the (pre)history of language(s) has, precisely since the advent of Chomsky's view of acquisition, been regarded as theoretically insignificant (because, so the argument goes, the acquisition device in the infant's brain has no knowledge of (pre)history). Once we realize that the mechanisms we must be concerned with are not just those of individual acquisition but also those whereby the behavior came into existence in the first place, this view is immediately exploded. The study of how various behaviors emerged in human prehistory becomes a crucial part of the effort to understand the innate mechanisms ultimately responsible for these behaviors. Much of this is considerably clearer when we think of a behavior like the kiss than in a case like language, precisely because the former appears not to be universal, so that its actual emergence in human culture is immediately seen as requiring some special circumstances. Oligogenesis may seem like a completely unrealistic, purely theoretical concept when we think of language, although this may be just because no one has really thought carefully enough about the problems and possibilities of the origin of language. There do seem to be other human behaviors which demonstrably originated a (very) small number of times but more than once, such as writing. Manaster Ramer (1989) also argues that there are aspects of language where this has happened (notably, the use of clicks as consonants, which is a nice coincidence given that the lip-kiss is fundamentally nothing but a bilabial click).

The lip-kiss may well be an example of oligogenesis, if, that is, we can substantiate its indigenous occurrence in some of the more remote places where this has been indicated. To be sure, the claims, based on interpretations of pottery or other art work, that the kiss was known in pre-Columbian America (e.g., Wundt 1910:135, Eibl-Eibesfeldt 1989: 243, who seems to be reading too much into Kauffman-Dog 1979) do not seem compelling for the most part. However, there is a tantalizing report by a reliable observer (Haddon 1890:336) of lip-kissing in yet another entirely different culture area, viz., among the Western islanders of Torres Straits (between Australia and New Guinea). If

Haddon was right that it was an indigenous custom to use "this salutation, combined with embracing the head, ... after a long separation, especially if the man had been supposed to be dead....", then it would be hard to believe that this could have come from Mesopotamia. Thus, oligogenesis seems at the moment to be quite likely in the case of the lip-kiss (as probably too in the case of writing, rhyme, click consonants, and so on). Whether the same is true of language is a question that needs to be faced, too, of course. The linguists' faith in strict monogenesis and extreme polygenesis as the only conceivable alternatives seems to be misplaced.

In any case, the little that we can glimpse of the prehistory of the lip-kiss suggests that it is one of the prime examples of why the traditional preoccupations with monogenesis (or the contrary) and innateness (or its opposite) are somewhat beside the point, and why as a result they have proved so barren of useful results. All forms of human behavior, unless completely instinctual, must be viewed as involving innate mechanisms which, under suitable circumstances, learn to reproduce such behavior from those around one, but also other innate mechanisms which, again only under the right conditions, can produce such behavior without any role models. To understand how any of these mechanisms function, or how they themselves evolved, it is essential to go beyond the usual dichotomies and to strive for concrete information about particular behaviors. The contrast between one origin and more than one, on this view, is meaningless, since the difference between one and two (or three or some other small number) is most likely to be an accident (or to be within the margin of error of our data). On the other hand, the difference between few (oligogenesis) and many (polygenesis in the narrower sense) may be significant, since the fact of oligogenesis would tend to suggest that we are dealing with a behavior which cannot arise very readily, i.e., one which requires a very particular kind of external circumstances (which in turn suggests that the innate mechanisms which respond to these circumstances are likely to have a very particular structure). Of course, we are still dealing just with indications and suggestions. In all such cases, certainly that of the lip-kiss, the real work remains to be done. But at least we have some new questions and some tried-and-true research methods to do the work with. This, too, seems to be very close to the situation in comparative linguistics.

Finally, let us conclude with yet another parallel between kissing and language. Although there was a time when the lip-kiss was sung as a sign of "civilization" (Lombroso 1893), its distribution seems to have nothing to do with the technological (or any other measure of) "level" or "degree of complexity" of a culture. Instead, it involves simple geography. Much as in every area of linguistics (and in the classic ethnological work on kinship systems and ideology by Todd (1983, and especially Sagart and Todd 1992), the key factor is propinquity (as one might well expect with kissing). For example, there is, or once was, a line (much like a linguist's isogloss) cutting across the continent of Eurasia, (and running somewhere through Southeast Asia in particular), northwest of which everybody (except the inhabitants of the

circumpolar regions) enjoyed the lip-kiss, but southeast of which no one did. We are almost certainly entitled to apply to the lip-kiss Sapir's (1921:219) famous words about language: When it comes to the kiss, too, "Plato walks with the Macedonian swineherd, Confucius with the head-hunting savage of Assam".

### References

- Andree, Richard  
1889 *Ethnographische Parallelen und Vergleiche*. Neue Folge. Leipzig: Veit.
- Barth, J.  
1893 *Etymologische Studien*. Leipzig: J. C. Hinrichs.
- Bloomfield, Leonard  
1933 *Language*. New York: Holt, Rinehart and Winston.
- Cohen, Jeffrey M.  
1982 "An Unrecognized Connotation of *nšq peh* with Special Reference to Three Biblical Occurrences". *Vetus Testamentum* 32, 4:416-424.
- Cooper, J. S.  
1983 "Küss". In: *Reallexicon der Assyriologie und vorderasiatischen Archäologie*, v. 6, pt. 5/6: 375-379.
- Darwin, Charles  
1871 *The Descent of Man and Selection in Relation to Sex*. London: J. Murray.  
1872 *The Expression of the Emotions in Man and Animals*. London: J. Murray.
- Dostal, Walter  
1983 *Egalität und Klassengesellschaft in Südarabien*. Horn/Vienna: Ferdinand Berger [Wiener Beiträge zur Kulturgeschichte und Linguistik, Veröffentlichungen des Instituts für Völkerkunde der Universität Wien, 20].
- d'Enjoy, Paul  
1897 "Le baiser en Europe et en Chine". *Bulletins de la Société d'anthropologie de Paris* 8: 181-185.



Eibl-Eibesfeldt, Irenäus

- 1979 "Human Ethology: Concepts and Implications for the Sciences of Man". *Behavioral and Brain Sciences* 2:1-57.  
 1989 *Human Ethology*. New York: Aldine de Gruyter.

Eichner, Heiner

- 1988 "Anatolisch und Trilaryngalismus". *Die Laryngalthorie und die Rekonstruktion des indogermanischen Laut- und Formensystems*, ed. Alfred Bammesberger, 123-151. Heidelberg : C. Winter.

Ellis, Havelock

- 1922 *Studies in the Psychology of Sex*, IV: Sexual Selection in Man. Philadelphia: F. A. Davis Company.

Gesenius, Wilhelm

- 1859 *Gesenius' Hebrew and Chaldee lexicon to the Old Testament Scriptures*, translated with additions and corrections from the author's Thesaurus and other works by Samuel Prideaux Tregelles. London : S. Bagster. [Translation of the author's *Lexicon manuale Hebraicum et Chaldaicum in Veteris Testamenti libros*, a Latin version of the work first published in 1810-1812 under title: *Hebräisch-deutsches Handwörterbuch des Alten Testaments*.]

Greenberg, Joseph H.

- 1960 "A Survey of African Prosodic Systems". *Culture in History*, ed. Stanley Diamond, 925-950. New York: Columbia University Press.

Gulik, Robert Hans van

- 1961 *Sexual Life in Ancient China: A Preliminary Survey of Chinese Sex and Society from ca.1500 B.C. till 1644 A.D.* Leiden, E.J. Brill.

Haddon, A. C.

- 1890 *The Ethnography of the Western Tribe of Torres Straits*. London: Harrison & Sons.

Hearn, Lafcadio

- 1896 *Out of the East*. Boston/New York: Houghton, Mifflin.

Hopkins, E. W.

- 1907 "The sniff-kiss in Ancient India". *Journal of the American Oriental Society* 28: 120-134.

Illich-Svitych, V[ladimir] M.

- 1964 "Drevnejšie indoevropejsko-semitskie jazykovye kontakty".  
In: V. N. Toporov (ed.), *Problemy indoevropejskogo jazykoznanija, Nauchnaja sessija, Tezisy dokladov*, 3-12. Moscow.

Kauffmann Doig, Federico

- 1979 *Sexual Behaviour in Ancient Peru*. Lima, Surquillo: Compactos.

Köhler, Ludwig and Walter Baumgartner

- 1948-53 *Lexicon in Veteris Testamenti libros*. Leiden: E. J. Brill.

Krauss, Friedrich S.

- 1911 *Das Geschlechtsleben in Glauben, Sitte, Brauch und Gewohnheitrecht der Japaner*. Leipzig: Ethnologischer Verlag. 2d ed.

Krauss, Friedrich S., and Tamio Satow

- 1965 *Japanisches Geschlechtsleben*. Hanau/M.: Schustek.

Leslau, Wolf

- 1987 *Comparative Dictionary of Ge'ez*. Wiesbaden: Otto Harrassowitz.

Longstreet, Stephen and Ethel

- 1970 *Yoshiwara: City of the Senses*. New York: McKay.

Lagarde, Paul de

- 1887 *Novae Psalterii Graeci editionis specimen*. Göttingen: Dieterich.

Lombroso, Cesare

- 1893 "L'origine du baiser". *Nouvelle revue* 83: 673-680.

Manaster Ramer, Alexis

- 1989 "Language, History, and Computation". In *Reconstructing Languages and Cultures* (ed. Vitaly Shevoroshkin), 104-109, Bochum: Studienverlag Brockmeyer.

To appear a "The Uses and Abuses of Mathematics in Linguistics". *Reports, Grup de Recerca en Lingüística Matemàtica i Enginyeria del Llenguatge (GRLMC)*. Tarragona, Spain: Universitat Rovira i Virgili.

To appear b "A Word to the Wise: Tubatulabal ooli- 'to get up'". In: Gene Casad (ed.), [A collection in honor of Wick Miller].

University of Utah Press.

- Meissner, Bruno von  
1934 "Der Kuß im alten Orient". *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Philosophisch-Historische Klasse* 28: 914-930.
- Melchert, H. Craig  
1994 *Anatolian Historical Phonology*. Amsterdam, Atlanta: Rodopi.
- Orbel, Margaret  
1985 *The Natural World of the Maori*. Auckland: Collins.
- Pokorny, Julius.  
1959 *Indogermanisches etymologisches Wörterbuch*. Bern: A. Francke AG.
- Puku'i, Mary Kawena.  
1972 "The Polynesian Family System in Ka'u, Hawaii." Rutland, VT: Charles E. Tuttle Co.
- Sapir, Edward  
1921 *Language*. New York: Hartcourt, Brace & World.  
1936 "Internal Linguistic Evidence Suggestive of the Northern Origin of the Navaho". *American Anthropologist* 38: 224-235.
- Sagart, Laurent, and Emmanuel Todd  
1992 "Une hypothèse sur l'origine du système familial communautaire". *Diogenes* 160:145-175.
- Schrijver, Peter  
1991 *The Reflexes of the Proto-Indo-European Laryngeals in Latin*. Amsterdam, Atlanta: Rodopi.
- Speiser, E. A.  
1964 *The Anchor Bible, Genesis: Introduction, Translation and Notes*. Garden City, NY: Doubleday.
- Todd, Emmanuel  
1983 *La Troisième planète*. Paris: Le Seuil.
- Wundt, Wilhelm  
1910 *Völkerpsychologie*, 4: Mythos und Religion. 2d. ed. Vienna: Wilhelm Engelmann.

## Relative Clauses in Eastern Shina\*

Peter Edwin Hook  
*University of Michigan*

In a pair of papers written almost twenty years ago E. Keenan and B. Comrie reported on a cross-linguistic study of relative clauses in which they found that the relativizability of positions in relative clauses can be formulated as a hierarchy (Keenan and Comrie 1977 and 1979):

- (1) subject > direct object > indirect object (>) oblique object >  
possessor > compared object

As later work on brain activation supports (Just et al. 1996), positions to the right on this hierarchy are progressively more difficult to relativize on and, consequently, more rarely encountered cross-linguistically. In general, if a language disallows relativization onto some position on the left of the hierarchy, say, indirect objects, then it also disallows relativization onto positions that are further to the right, say, possessors and objects of comparison.

Prenominal relativization strategies in general are less explicit than postnominal ones. This is because prenominal relatives commonly if not invariably involve the gapping, deletion, or absence in the relative clause of any token of the noun phrase shared with the matrix clause. Together with this gapped noun phrase also absent is the postposition or particle or case affix that indicates the relation of the referent of the gapped noun phrase to the action expressed by the verb in the relative clause. Consequently there is a much higher likelihood for ambiguity or indeterminacy in the interpretation of prenominal relative clauses than in the interpretation of their postnominal counterparts.

Two prenominal strategies can be elicited from Shina speakers: The more explicit one uses a form of the interrogative or indefinite pronoun (which in E. Shina are the same) to represent the shared noun phrase inside the relative clause:

- (2) [kesi<sub>i</sub>      myei madat thaw] (Zu<sub>i</sub> kone    gaw)<sup>1</sup>  
       who.Erg my    help    did      that where went

‘(Where did the one [who helped me] go)?’ (1989 field notes)

In the other strategy the shared noun phrase is ‘gapped’ inside the relative clause which itself, with the addition of the suffix *-(e)k*, becomes a noun phrase in the matrix:

- (3) [(e)<sub>i</sub> myei madat thaaw]-ek<sub>i</sub> kone    gaw  
       (gap) my    help    did-one    where went

‘Where did (the one [*who* helped me]) go?’ (1989 field notes)

In situations where the case affix is zero the addition of *-(e)k* is optional [see also exx. (7), (11), (15), (18)]:

- (4) musu [(e)<sub>i</sub> goZ-e    hAI] karaar<sub>i</sub>    lishaar-emis  
       I.Erg (gap) house-Obl is.Fsg knife(Fsg) hide-1FsgFut

‘I’ll hide (the knife [*which* is in the house]).’ (History of Astor)

Since the first strategy is one in which less information is lost one might expect to find it used especially when speakers relativize onto less accessible positions. However, an examination of texts reveals that despite its greater explicitness the first strategy is never used in natural discourse, not even when use of the second strategy leads to the loss of so much information that interpretation outside of a situational context is impossible. The first strategy, evidently an artefact of the use of Urdu to elicit data, is a calque on the relative-corerelative construction found in that language.

Examples from texts that illustrate the progressively less accessible points on the Noun Phrase Hierarchy follow:

---

<sup>1</sup>The bracketing of example (2) assumes that neither clause in a relative-corerelative construction is embedded in the other (see E. Keenan 1985:164ff). Since in E. Shina relative-corerelative clauses do not appear in natural data, making such an assumption has no consequences for the discussion that follows.

## I. Relativization on subjects:

- (5) (Zi [Sac-íí hAA]-k) phat b-il-e  
 those stick-Ger are-ones loose . become-Pst-3pl  
 ‘(The ones [*who* were stuck together]) were released.’  
 (R&P<sup>2</sup> 48)
- (6) “([anu ashup-e koi-se paNyo b.il.o]-k)-are  
 this horse-Acc any-Erg mount became-one-Dat  
 kacaak-ek inaam d-on.”  
 so.much-one prize give-1plFut  
 “‘We shall give a prize to (the one [*whoever* can mount this  
 horse]).’” (Kesar 135)

## II. Relativization on objects:

- (7) Zu-se dady-ere ([jol-ejaa wyaw] bai) khal-eé daw  
 he-Erg old.woman-Dat sack-Loc put.3sg food take.out-CP gave  
 ‘Taking out (the food [*which* he had put in the sack]) he gave it  
 to the old woman.’ (Maamad Sher Ali)
- (8) ([tu-se khyUU hAAw]-ek) ([mu-su khyUU hAAUs]-ek)-ejo  
 you-Erg eating are-one I-Erg eating am-one-Abl  
 so hAU  
 good is  
 ‘(That [*which* you are eating]) is better than (that [*which* I am  
 eating]).’ (1994 field notes)

## III. Relativization on indirect objects and dative possessors:

- (9) ([tu-se rupaaye daa]-k baal) kone hAU  
 you-Erg rupees gave.2sg-one boy where is  
 ‘Where is (the boy [*to whom* you gave money])? (1994 notes)

---

<sup>2</sup>“R & P” refers to a line in the story of the queen and the bald man (*roNi gaa phaRaaro*). A number preceded by “Kesar” refers to a line in the story of Kesar of Layul (see Hook 1996). Both were narrated to me in Skardu in October 1989 by Mr. Nasir Hussain. “Proverb”, “Maamad Sher Ali”, “Dudusher Gaav”, and “History of Astor” are unpublished materials collected by Nasir Hussain and transcribed by me in 1994.

- (10) ([shakal nush]-ek)-ere akal nush  
 beauty is.not-one-Dat wit is.not  
 ‘(The one [*who* has little beauty]) has little wit.’ (Proverb 70)

#### IV. Relativization on oblique objects:

- (11) “([baks-ijo pUil-e maar-aas] karaar) khal-e!”  
 box-Abl Pu~ilo-Acc kill-1MsgPst knife(Fsg) take.out-Imper  
 “‘Take (the knife [*with which* I killed Pu~ilo]) out of the box.”  
 (History of Astor 5.8)
- (12) ([Zo i beey]-ek jery)-e gin-i wazh-oni razh-e  
 he self sits-one old.F-Acc take-CP come-Inf say-3plPst  
 ‘They told him to bring (the old woman [*with whom* he sits]).’  
 (R&P, line 48)
- (13) ([ikbaal-ere samaan lad-u]-k dukaan) kone hAI?  
 Iqbal-Dat supplies(Fsg) got-Msg-one store(Fsg) where is.Fsg  
 ‘Where’s (the shop [*in which* Iqbal got supplies])?’ (1994 notes)<sup>3</sup>

#### V. Relativization on possessors:

- (14) ([sAIyO nush]-ek)-i hat doyou thii  
 rations be.not-one-Erg hands washed they.say  
 ‘(The one [*who* has no rations]) washed his hands...’  
 (Proverb 72 [referring to a moocher preparing to eat])
- (15) ([nau pangave hAU] Ashup) gaa krino  
 nine stirrups is horse also rotted  
 ‘(The horse [*who* has nine stirrups]) also rotted.’  
 (Maamad Sher Ali 1994:33)

<sup>3</sup>Notice that the verb form *ladu-* ‘got’ is masculine singular agree-ing in gender and number with the dative noun phrase *ikbaal-ere* ‘to Iqbal’. Agreement with “dative subjects” is a peculiarity of the grammar of the verb *laj-* ‘get’ and of some predicates of experience in Gultari and other easterly dialects of Shina. See Hook 1990 and Hook 1996:172-4.

Notice that in examples (14) and (15) the shared noun phrase may be regarded as the 'logical' subject of the predicate of possession in the relative clause. That is, the possessor is part of the argument structure of that predicate. In the natural (unelicited) data the closest I have to an example in which the possessor noun phrase modifies some other noun phrase in the embedded clause is (16):

- (16) ([karkaT-ijaa shaS hAAw]-ek) tagaal-o  
 Karkat-Loc mother-in-law(Fsg) is.Msg-one lucky-Msg  
 'Lucky is (the one [*whose* mother-in-law is in Karkat]).'  
 (Proverb 70)

However, elicited examples of gapped possessors that function as modifiers of noun phrases appear below in exx (19) and (32-3).

#### VI. Relativization on indeterminate positions:

- (17) ([shU nush]-ek shafat)-ejaa gwake ne dyaa  
 dog be.not-one dish-Loc fight don't give  
 'Don't fight over (a dogdish [*for which* there is no dog])!' (Proverb 85)
- (18) ([kuNo nush] aSe)  
 corpse be.not tears (Proverb 128)  
 '(Tears [*which* no-one has died to cause])!' (=crocodile tears)

It is possible to elicit examples in which the loss of information is sufficient to render the result ambiguous, at least when presented out of context:

- (19) ([baal-i cori thaaw]-ek)-i ripoT ne daw  
 boy-Erg robbery did-one-Erg report not gave  
 '(The person [*from whom* the boy stole]) didn't report it.'  
 '(The person [*whose* boy stole]) didn't report it.' (1994 notes)

The ability to relativize on possessors is not so common in languages using prenominal relativization strategies. According to Comrie (MSS), for example, it is not possible in the Northeast Caucasian language Tsez in situations involving alienable possession:



- (20) \*? g<sup>w</sup>aj b-oxi-n b-âk'i-ru uzhi ijaj-xo  
 dog An-run-Ger An-GO-PstP boy cry-Pres  
 'The boy *whose* dog has run away is crying.'

In Marathi, however, relativization on possessors is commonly seen although my impression is that, like in E. Shina, its occurrence is more common in situations of inalienable possession<sup>4</sup>, as in (21):

- (21) ([mula as-l-el]-i loka) nehama dukhi dis-t-aat kaa?  
 kids be-Pst-PP-Npl people always unhappy look-Pres-3pl QM  
 'Do (people [*who have* children]) always look unhappy?'

There are two questions that I pursue in this exploratory study of the grammatical properties of the prenominal relative construction in the Gultari dialect of Eastern Shina: 1. Is the shared noun phrase invariably gapped? 2. Is the relative clause really a clause?

<sup>4</sup>Even the expression of a kind of alienable possession via the pre-nominal strategy is not impossible in Marathi if some strandable locative postposition (like *dzaval* 'near') remains in the clause:

- (a) ([tikiT dzaval n-as-l-el]-yaa lokaa-ni) hyaa raange-t  
 ticket near not-be-Pst-P-Obl people-Erg this queue-Loc  
 ubha raahu naye-t  
 standing stay shouldn't-pl

'(People [who don't have a ticket]) should not stand in this line.'

However, the prenominal strategy appears not to be open to the expression of alienable possession if a locative postposition used in expressions of possession is one that cannot be stranded:

- (b) \*([tikiT kaDe n-as-l-el]-yaa lokaa-ni) hyaa raange-t  
 ticket near not-be-Pst-P-Obl people-Erg this queue-Loc  
 ubha raahu naye-t  
 standing stay shouldn't-pl

'(People [who don't have a ticket]) should not stand in this line.'

Another way around the infelicity of using prenominals in the expression of alienable possession is to shift its temporary component onto some other locative relationship in the relative clause:

- (c) ([paise khiShyaat n-as-l-el]-yaa lokaa-ni) philim  
 money pocket.Loc not-be-Pst-P-Obl people-Erg film  
 paah-aay-laa dzaa-u naye-t  
 see-Inf-Dat go-Inf shouldn't-pl

'(People [who don't have money in their pockets]) should not go to see movies.'

I am grateful to Madhav and Shubhangi Deshpande for checking some of these examples and suggesting others.

1. Is the shared noun phrase invariably gapped? Notice that the shared noun phrase in ex. (6) is present inside the relative clause:

- (6) "[anu ashup-e **koi-se** paNyo b.il.o]-k)-are  
       this horse-Acc any-Erg mount became-one-Dat  
       kacaak-ek inaam d-on."  
       so.much-one prize give-1plFut

"We shall give a prize to (the one [**whoever** can mount this horse])."

(Kesar 135)

Examples of this kind appear to be limited to cases when the speaker does not presuppose the specific identity or even the existence of any referent matching the characterization spelled out in the relative clause:

- (22) ([anu pezaar **kes-ere** gaa kar bilo]-k)-esi.naalaai kash th-emus  
       this slipper who-Dat also fit became-one-with marriage do-1sgPr  
       'I'll marry (the one [**whomever** this slipper fits]).' (1994 notes)

It appears that there is another kind of relative clause in E. Shina in which the shared noun phrase is not gapped and [unlike in (6) and (22)] the specific identity or even existence of a referent that matches the characterization given in the relative clause is presupposed:

- (23) ([se cai tu i se traaye-re phal.th-aa]-k) mo bil.aas  
       that bird you Emp Erg window-Dat throw-2MsgPst-one I was

'I was the bird you tossed out the window.' (Dudusher Gaav)

At first examples like (23) may look like instances of prenominal relative clauses extraposed to the right of their head nouns (as occur in Basque: de Rijk 1972). If so (23) should be bracketed differently:

- (23') (se cai), ([tu i se traaye-re phal.th-aa]-k), mo bil.aas  
       'That bird, the one you tossed out the window, was me.'

However, this seems an unlikely analysis of the Shina: For one thing, the noun *cai* 'bird' is feminine, as is the speaker (*mo*), while the copula *bilaas* 'was' is a masculine form, in indirect agreement with the subject of the embedded clause. Leaving *cai* inside the relative clause and having the copula agree with the relative clause's nominalized predicate *phal.thaa-k* allows an explanation for the masculine suffix in *bil.aas*. Secondly, the elicitation of other examples reveals that the leftmost noun phrases in them do not behave like the

subjects of matrix copulas. Rather they get whatever case is required by the predicate in the embedded clause:

- (24) ([tu-re son lej-ony]-ek) nush  
you-Dat gold get-Inf-one not

'You are not about to get the gold!'<sup>5</sup> (1994 fieldnotes)

- (25) ([musu anu krom thy-ony]-ek) nush  
I.Erg this work do-Inf-one not

'I'm not about to do this job!' (1994 fieldnotes)

Should we conclude then that examples (23-25) are all instances of internally-headed relative clauses like those found in Diegueño<sup>6</sup>?

<sup>5</sup>Even the expression of a kind of alienable possession via the pre-nominal strategy is not impossible in Marathi if some stranda-ble locative postposition (like *dzavaL* 'near') remains in the clause:

- (a) ([tikiT dzavaL n-as-l-el]-yaa lokaa-ni) hyaa raange-t  
ticket near not-be-Pst-P-Obl people-Erg this queue-Loc  
ubha raahu naye-t  
standing stay shouldn't-pl

'(People [who don't have a ticket]) should not stand in this line.'

However, the prenominal strategy appears not to be open to the expression of alienable possession if a locative postposition used in expressions of possession is one that cannot be stranded:

- (b) \*([tikiT kaDe n-as-l-el]-yaa lokaa-ni) hyaa raange-t  
ticket near not-be-Pst-P-Obl people-Erg this queue-Loc  
ubha raahu naye-t  
standing stay shouldn't-pl

'(People [who don't have a ticket]) should not stand in this line.'

Another way around the infelicity of using prenominals in the ex-pression of alienable possession is to shift its temporary compo-nent onto some other locative relationship in the relative clause:

- (c) ([paise khiShyaat n-as-l-el]-yaa lokaa-ni) philim  
money pocket.Loc not-be-Pst-P-Obl people-Erg film  
paah-aay-laa dzaa-u naye-t  
see-Inf-Dat go-Inf shouldn't-pl

'(People [who don't have money in their pockets]) should not go to see movies.'

I am grateful to Madhav and Shubhangi Deshpande for checking some of these examples and suggesting others.

<sup>6</sup>In their discussions of internally headed relative clauses the exam-ples that Comrie (1989:145) and Keenan (1985:162) adduce that seem most similar to those of E. Shina are those from Diegueño:

Perhaps not: The meaning of examples like (24-5) is reported to involve an emphatic future (as reflected in their English translations) rather than the restriction on the domain of a noun phrase's referents that one would expect from a relative clause. Moreover, the putative relative clauses in some of them allow the occurrence of the topic particle *to*:

- (26) ([? mo to ([? aanaa-yo khar.the waapas waj-ony]-ek) nush  
I.Nom Top here-from downward back descend-Inf-one not

'As for me, I'm not about to go back down from here!'

(Maamad Sher Ali)

Given the bar on the occurrence of noun phrases marked with topic particles in Japanese and Korean relative clauses (Kuno 1973:254, Na 1986), it would be surprising to find topic particles having scope over noun phrases inside Shina relative clauses.

On the other hand, to argue that the leftmost noun phrases are not part of the embedded clauses in exx (24-25) would require us to posit a rule of case attraction to account for the ergative in (24) and the dative in (25). Such a rule is required to handle other phenomena in the grammar of Shina<sup>7</sup> and may well apply here, too.

- 
- (a) Tɔ̃nay ?ɔ̃wa: ?ɔ̃wu:w-pu-Ly ?ciyawx  
yesterday house see.Pst1sg-Def-Loc sing.Fut1sg  
'I will sing in the house that I saw yesterday.'

The Diegueño "definitizer" *-pu-*, directly affixed to the finite form of the verb while itself taking case affixes may be compared to the E. Shina *-(e)k-*, while the presence of a full noun as token of the shared noun (in this instance ?ɔ̃wa: 'house') can be compared to *cai* 'bird' if *cai* is indeed inside the relative clause in (23).

<sup>7</sup>The evidence in support of attraction comes from constructions composed of a conjunctive participial (CP) form followed by a form of the stative verb *as* 'be' and is discussed in Hook (1996: 181). In (a) the pronoun *Zise* 'him' is the subject of the stative *asil-o* 'was-3sgM', but owes its accusative case to its being the direct object of *ban th-* 'close up; shut in':

- (a) Zis-e kamaraa-k-ejaa ban th-eé as-il-o  
him-Acc room-one-Loc closed make-CP be-Pst-M3sg  
'He (Bubalastang) was closed up in a room.' (Kesar 45)

By making substitutions we can obtain evidence that the mascu-line singular form *asilo* 'was' in (a) is not a default form but one that shows concord with the accusative 'subject' *Zise*. Replacing *Zise* with the corresponding accusative plural form *Zino* forces the verb *asil-* 'was' to take the plural suffix *-e* to accord with it:

- (b) Zin-o kamaraa-k-ejaa ban th-eé as-il-e  
them-Acc room-one-Loc closed mak-CP be-Pst-M3pl  
'They were closed up in a room.'

2. Is the relative clause really a clause? Given examples like (11) in which the embedded predicate form *maaraas* '(I) killed' is fully specified for tense, person, number, and gender, one is apt to conclude that (leaving aside the gapping of the shared noun phrase) the E. Shina relative clause is a full clause:

- (11) "([baks-ijo pUil-e maar-aas] karaar) khal-e!"  
 box-Abl Pu'ilo-Acc kill-1MsgPst knife(Fsg) take.out-Imper  
 "Take (the knife [with which I killed Pu'ilo]) out of the box."  
 (History of Astor 5.8)

But active manipulation of such examples proves otherwise. Notice that in (27) the predicate form *khatu* 'emerged' is a masculine singular, in apparent agreement with the embedded subject *nom* 'name':

- (27) ([nom 'ne khat-u]-k)-i khei chiny-aw  
 name not emerged-Msg-one-ErgMsg bridge broke-3Msg  
 '(He [whose name had not emerged]) broke a bridge.'  
 (Proverb 129)

However, if the breaker of the bridge is not male or not singular, the masculine singular form *khatu* is not accepted. Rather, the past tense form of the intransitive predicate *khaj* 'emerge; climb' agrees in gender and number not with its subject *nom* 'name' but with the noun modified by the relative clause (exx from 1994 field notes):

- (28) ([nom 'ne khat-y]-ek)-o khei chiny-ei  
 name not emerged-Fsg-one-ErgFsg bridge broke-3Fsg  
 '(She [whose name had not emerged]) broke a bridge.'

---

Although there is a conjunctive participial form *ban thee* 'having closed' in (a) and (b), it does not have the function of conjunction. Rather, it expresses a state. Compare (a) with (a'):

- (a') \*? (koi-se) Zis-e<sub>i</sub> kamaraa-k-ejaa ban thaw tato Zo<sub>i</sub> asil.o  
 someone-Erg him-Acc room-a-Loc shut did then he.Nom was  
 \*?'Someone closed him up in a room and then he was.'

Since speakers are reluctant to accept (a'), and even when they do accept it, deny that it is a paraphrase of (a), we cannot regard the conjunctive participial form in (a) and (b) as having a conjunctive function. The constructions in (a) and (b) are monoclausal and to account for the accusative case in their pronouns we have to posit a rule of case attraction.

- The morphological behavior of these examples is reminiscent of a class of compounds in Hindi-Urdu which contain intransitive past participles that agree not with their subjects but their head nouns:

- To add a further layer of complexity animacy may also have a role to play: If the embedded subject is animate the concordant parts of the predicate may agree with either it or the head noun:

- ‘Where is (the woman [*whose* children are many])?’

But in other instances involving animate subjects this choice is not possible. The embedded predicate must agree with the head noun and not with its subject (elicited data from 1994 field notes):

- (33) ([baal uCit-y-ek)-o                      ripoT   dyei  
       boy ran.away-Fsg-one-ErgFsg   report   gave.3Fsg  
       ‘(She [*whose* boy ran away]) made a report.’

Perhaps a morphological explanation is possible in which the relative clause's predicate together with its suffix *-(e)k* and case affix would form a single word which must satisfy certain well-formedness conditions. One of these conditions would rule out discordant stacking of agreement affixes: ie, \*words that are simultaneously marked for two different genders and/or numbers. Since *exx* (27-9)

and (32-3) all have head nouns with ergative affixes that distinguish the number as well as (in the singular) the gender of their referents, these nouns cannot also agree with the subjects of their respective modifying clauses. (31), however, has an embedded predicate with only one slot for gender/number concord and may agree either with the subject of the relative clause or with the noun *cei* 'woman' that it modifies, without violating any such morphological constraint.

However plausible such a constraint may seem, it will probably have to be limited to situations in which the embedded subject is a third person. First person subjects allow the head noun to show two discordant number agreements [(34) is elicited data]:

- (34) ([musu cori th-aas]-ek)-ojaa ripoT ne dye  
 I.Erg robbery do-1MsgPst-one-Ergpl report not gave.Mpl  
 ' (The ones [from whom I stole]) did not report it.'

The morphological constraint, if it exists, appears to be interacting with hierarchies of animacy and person.

Whatever set of explanations proves to be optimal, we must conclude that what at first sight may appear to be fully fledged finite predicates in E. Shina relative clauses do not have the same properties that finite predicates in non-embedded clauses do. While the assignment of cases to a predicate's arguments is identical in both relative and independent clauses in Shina<sup>8</sup>, gender-number concord in relative clauses may be with the head noun rather than with the clausal subject. In this respect the relative clause of E. Shina is comparable with the relative participles found in Marathi, Gujarati, and other northern and western Indo-Aryan languages. However, it differs from relative participles in these and other Indo-Aryan languages in allowing agreement in person with the clausal subject.

<sup>8</sup> Contrast the use of the genitive (rather than the nominative or the ergative) for the embedded subject in the prenominal construction in Urdu (a) and Kashmiri [ex (b) is from Raina 1991:53]:

- (a) jab raam-ne shaam-kaa kiy-aa.huaa kaam dekhaa  
 when Ram-Erg Sham-Gen do-PstPmsg work(msg) saw  
 (b) yelyi raam-an shaam-sinz ker-mits keem vuch  
 when Ram-Erg Sham-Gen do-PstPfsg work(fsg) saw  
 'When Ram saw the work which Sham had done...'

\* The present version of this paper on the grammatical properties of an Indo-Aryan language spoken in Central Asia has been written to help celebrate the life and career of my colleague and friend, Professor Vitaly Shevoroshkin. An earlier incarnation was presented at the seventeenth meeting of the South Asian Languages Roundtable (SALA-17), held at the University of Texas (Austin), 2-4 June 1995. The research on which it is based was conducted in Skardu, Northern Pakistan, during the fall of 1989 with financial support from the Smithsonian Institution as part of the project on folk cultures of Pakistan organized by Wilma Heston and William Hanaway of the University of Pennsylvania in affiliation with Lok Virsa, Shakarparian, Islamabad, Pakistan; and during the spring of 1994 with financial support from the American Pakistan Research Organization (APRO) as part of the "Linkages Project" with the National Institute of Pakistan Studies (NIPS), Quaid-e-Azam University, Islamabad, Pakistan. For invaluable help in transcription I am indebted to Ruth L. Schmidt who visited Skardu in 1994. I am very grateful to Bernard Comrie for his perceptive comments on an earlier version of this, and I count it a great stroke of good fortune to have encountered Mr. Mohd. Nasir Hussain in the fall of 1989. Without him these pages could not have been written.

Transcription: Symbols used have the values normally found in descriptions of modern Indo-Aryan languages. I use doubling rather than macron or semi-colon to indicate length in vowels. While rising and falling tones exist in E. Shina, I did not attempt to record them in my transcription, except in gerunds. Cap forms of stops and fricatives / *T, Th, D, R, N, C, Ch, Z, S* / stand for retroflexed counterparts of sounds represented by non-caps. The digraph *sh* stands for palatal "esh"; while *ng* represents "engma". Nasal vowels are shown by the cap forms of the corresponding oral vowels. A following consonant tends to raise vowels; a preceding *Z*, to lower and centralize them. Intervocally the palatal affricate *j* is often realized as a fricative (*zh*). Abbreviations include:

Abl.....ablative	F.....feminine	N.....neuter
Acc.....accusative	Fut.....future	Obl.....oblique
An.....animal class	Ger.....gerund	Pr.....present
Caus.....causative	Imper.....imperative	Pst.....past
Dat.....dative	Inf.....infinitive	P.....participle
Def.....definitizer	Loc.....locative	QM..question marker
Erg.....ergative	M.....masculine	Top.....topic marker



## References

- Comrie, Bernard.  
 1989 *Language Universals and Linguistic Typology*. Univ. of Chicago Press.
- MSS Relative clauses and related constructions in Tsez. Paper presented at NSL-9, Univ. of Chicago, 1995.
- Corbett, Greville G.  
 1991 *Gender*. Cambridge University Press.
- de Rijk, R.P.G.  
 1972 Relative clauses in Basque: a guided tour. In P. M. Peranteau, J. Levi, G. Phares, eds. *The Chicago Which Hunt: Papers from the Relative Clause Festival*. Chicago Linguistics Society. Pp. 115-35.
- Hook, P.E.  
 1990 A Note on Expressions of Involuntary Experience in the Shina of Skardu. *Bulletin of the School of Oriental and African Studies* 53:77-82.  
 1996 Kesar of Layul: A Central Asian Epic in the Shina of Gultari. In *Studies in the Popular Culture of Pakistan*. Edited by William Hanaway and Wilma Heston. Lahore: Sang-e-Meel. Pp. 121-183.
- Just, M.A., P. Carpenter, T. Keller, W. Eddy, K. Thulborn  
 1996 Brain Activation Modulated by Sentence Comprehension. In *Science* 274:114-6.
- Keenan, Edward L.  
 1985 Relative clauses. In *Language Typology and Syntactic Description*, vol. 2. Edited by Timothy Shopen. Cambridge University Press. Pp. 141-70.
- \_\_\_\_\_ and Bernard Comrie  
 1977 Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63-99.  
 1979 Data on the noun phrase accessibility hierarchy. *Language* 55:333-51.

Kuno, S.

1973 *The Structure of the Japanese Language*. MIT Press.

Na, Y.

1986 *Syntactic and Semantic Interaction in Korean: Theme, Topic and Relative Clause*. Univ of Chicago PhD dissertation.

Raina, Achla Misri

1991 *An S-Selectional Approach to Grammar: Some Issues in Kashmiri Syntax*. Unpublished PhD diss. Kanpur: Indian Institute of Technology.

Sobati, Krishna

1972 *Daar se bichuRii (A Leaf Torn from the Branch)*, 2nd ed. Delhi: Rajkamal.

## LUVIAN COLLECTIVE AND NON-COLLECTIVE NEUTRAL NOUNS IN -AR

Vyacheslav Vs. Ivanov  
*University of California at Los Angeles*

Calvert Watkins was the first to establish the meaning of the Luwian substantive *wa-a-ar-ša* “water” (Watkins 1987; 1994:309-314; 1995:144-145). This discovery was linked to his new interpretation of the part of the Luwian ritual of Puriyanni KUB XXXV 54 III 12 ff. (the text of the beginning of the XIVth century, Starke 1985:55-71). In general one can accept his way to understand the text with a necessary change as far as the noun *ali-* is concerned. As it was suggested by Meriggi (1957:215; cf. Laroche 1959:26-27; Carruba 1982a:47-48) the latter means “sea”; this supposition has been proved by the analysis of the use of the word in a group of the Luwian rituals connected to the birth of a child (Starke 1985:205, 210, 215; 1990). It fits very well in the interpretation of the ritual of Puriyanni as the charm of water and salt given on the base of the Hittite introduction by Laroche and accepted by Watkins (Laroche 1959:152; Watkins 1995:144, cf. already Meriggi 1957:203, 215; Carruba 1982a:47-48). In this conjuration as the source of water the river is mentioned whereas that of salt is the sea rock: *[w]a-a-ar-ša-at-ta* *ÍD-ti* *[na-na-a]-am-ma-an* *[M]UN-ša-pa a-a-al-la-a-ti u-wa-a[-ni-ya-ti]* *ú-pa-am-ma-an* *[w]a-a-ar-ša-at-ta zi-i-l[a ÍD-i]* *an-da* *[n]a-a-wa i-ti MUN-ša-pa-a[t-ta z]i-la* *[a-a]-li-i u-wa-a-ni-ya na-a[-wa i-t]i* “This is the [w]ater [got] (= [driv]en) from the river, and this is the [s]alt brought from the sea ro[ck]; the [w]ater will never go to th[e river] and the salt will nev[er g]o to the [se]a rock” (KUB XXXV 54 Rs. III 17-21 = II.1 according to the scheme of Starke 1985:55-59, 68-69; the last two sentences are repeated with small variations in a fragment KUB XXXV 47 2'-5' dated by the XIIIth century: Starke 1985:58-59, 71 = III.2). The use of the case ending of the “animated neuter” *-ša* (Carruba 1982; 1992; van den Hout 1984; see already Bajun 1978) in the two last sentences of the text might have been motivated by the active function of the verb *i-* “to go” (cf. Ivanov 1981), with which the nouns *war-ša* “water”, *MUN-ša* “salt” are connected as grammatical subjects. In two preceding sentences this animated function of the two nouns has been anticipated. That may explain the use of the same case endings in the constructions in which these substantives precede the forms of the nominative-accusative neuter of the mediopassive participles in *-(a)mman*: *upa-mm-an* “brought” and (preserved only in its ending) *[nanna]-mm-an* “[driv]en, [le]d”, see on the other details of the syntax of the passage: Meriggi 1957:204; Carruba 1982a:47-48. The conclusion that at least in this text the case is that of the animated neuter rather than a quasi-ergative can be confirmed by the continuation where it is said that *[wa]-a-ar-ša ...* *ḫa-[-la]-a-al* (ib., Rs. III 25) “the water (is) pure” (Laroche 1959:152; literally: purity, cf. Carruba 1982:15; a shortened

adjective according to Meriggi 1980:282): here a substantive with the same case ending is connected to the neuter noun used in a predicative function.

The stem *wa-a-ar-* “water” clearly can be traced back, as Watkins has remarked, to the Indo-European *\*wo:r-* “water”: Old Indian *va:r*, *va:ri* “water”, Shina *ba:ri* “small lake”, Sinhalese *vāralla* “light rain” (Turner 1989:674, N 11556), Avestan *va:r* “rain”, *vairi/varay-* “lake” (Bartholomae 1979:1364-1365, 1410-1411), Old Armenian *gayr* “swamp”, Tokharian A *wär*, B *war* “water”, Old Norse *úr* “drizzle”, *ver* “sea” (poetical), Old English *waer* “sea” (rare), Lithuanian *jūra* “sea”, Latvian *jūra* “sea”, Prussian *iu:rin* “sea” (Toporov 1975:160-161; Pokorny 1959:78; Gamkrelidze, Ivanov 1995:I, 580). The correspondence to the long vowel in Old Indian makes it possible to reconstruct length in Indo-European, cf. Melchert 1994:245,265 (on Proto-Anatolian). Yet another form of the same root with a different structure is also represented in Luvian.

In the Cuneiform Luvian text of an incantation of the beginning of the XIV th century B.C. KUB XXXV 107 +108 Rs. III 19' (see on the text in which the mythological parts are mixed up with the rituals: Starke 1985, 210 ff., 238) the expression IGI.<sup>HLA</sup>-*wa a-an-da ú-wa-ar-ša* was interpreted by P.Meriggi (1957:215) and following him recently by F.Starke (1990:328-329) as describing “tears” (*uwar-ša*) that are wiped out of the “eyes” (IGI.<sup>HLA</sup>). This interpretation of the construction is based on the analogy with the Hittite-Luvian medicine text of the XIII century B.C. KUB VIII 38+ XLIV 62 Rs. III 20' ff. : *nam-ma-an a-an-da-az A-az [i]š-ḫa-aḫ-ru ši-pa-an(-)x [...ar-ha a-an-aš-zi]* “wischt er ihn mit warmem Wasser, [und zwar] die Tränen und den... weg” (Burde 1974:31). The validity of the parallel is strengthened by the numerous traces of the Luvian influence in this medical text just in those parts of it that describe eyes (ib., Rs. 12': Luw. *ta-a-u-i-iš-ši* “his eye”, Laroche 1959:96; Burde 1974:34; 75; Starke 1990:131, n. 394) and “tears” (ib., 10': Hit. *iš-ḫa-aḫ-ru* with the Luvian attribute *i-ya-[u-wa]-an*, used with a glossal cuneiform sign, “Glossenkeil” in the text of the XIIIth century KUB XXX 33 I 9, Laroche 1959:51; Burde 1974:33-34; Starke 1990:35). The mythological story included into the Luvian incantation describes the feast arranged for (all) the gods( (Luvian DINGIR<sup>MES</sup>-. *in-zi pu-u-na-ti-in-za*; Ivanov 1996:719: a Greek-Tokharian-Luvian isogloss in the expression of totality possibly linked to the numeral “fifth”, Lycian *pīnuta-*, cf. Shevoroshkin 1979:188 ). An interesting semantic parallel found in Luvian *maššaninzi punatinza* = Hittite *ḫumanteš šiuneš* = Vedic *viśvádeva:s* (for the Vedic hymns on these gods see a survey: Renou 1958) makes it possible to suggest a reconstruction of an Indo-European mythological and ritual prototext containing or describing the invitation of “all the gods” (including deified mountains, streams and the Ocean in all these traditions) for a feast. According to the Hittite and Luvian versions different gods had been invited. But the god of the eye diseases had not been among the guests and was offended. In this way the

notion of the eye disease and the magical means to cure it are introduced in the text of the incantation.

There are semantic and syntactical difficulties in this Luwian fragment that have not been discussed in the previous studies. The tears should be wiped out (Hit. *arḥa*) of the eyes, but the adverb *anda* that is rightly seen here by F. Starke usually means "in" and not "out". Starke remarks: "da das Adverb auch lokativische Funktion hat, widerspricht es dieser Auffassung nicht" (Starke 1990:329, n. 1164), but the meaning of *anda* "in" is confirmed by many Luwian contexts (cf., for instance, Laroche 1959:29).

The Luwian case of the animated neuter in *-ša uwar-ša* (Carruba 1982:5) suggests an active or instrumental meaning (either of a subject or of an object). In the Hittite-Luwian medicine text cited above the Hittite (*a-a-an-da-az*) *A-az* "with warm water" seems to be comparable to Luw. (*a-an-da*) *u-wa-ar-ša* in the incantation. The equation

$$\text{Hit. } A\text{-}az = \text{Luw. } u\text{-}wa\text{-}ar\text{-}ša$$

is grammatically valid since the Luwian case in *-ša* (*/-za*) corresponds in its function to the Hittite instrumental form in *-a(n)z* (cf. on the possible link of this form and the Hittite quasi-ergative: Garrett 1990; Dixon 1994:187-188; Carruba 1992). It seems possible to suggest a reinterpretation of the incantation based on the proposed meaning of the Luwian word and its case form of the animated neuter. The Luw. *uwar-ša* "by the water" is used in the construction *a-wa-at-ta* IGI.<sup>HLA</sup> *wa a-an-da ú-wa-ar-ša lu-u-wa-an-da*, KUB XXXV 107 + 108 Rs. III 19' = IIa) III.1 according to Starke 1985:238. The Luwian verb *luwa-* "to pour" (Melchert 1988:217; 1993; 1994:72-73, 238, 241, 262; cf. Starke 1990:224, 327-328, 378, 455-456) in the form of 3 Person Plural Past agrees with the subject IGI.<sup>HLA</sup> *wa* (= *tawa*). The combination of the verb with the name of "water" as an object is quite similar to the corresponding Hittite construction (*watar* + *lahḫuwai-*) attested in a number of texts (Güterbock, Hoffner 1980:14); Hittite parallels abound also for a combination with the preverb *anda* (ib.). Thus it seems possible to suggest a translation "and it was told (Luw. *a-wa-at-ta*, Laroche 1959:21) that the eyes poured down the water (=the eyes were filled with tears)". (The description of tears as "water of an eye" is a linguistic universal.)

According to this interpretation of the incantation, in it the Luwian word for "water" has the form *ú-wa-ar-ša* which differs from the one discovered by Watkins (*wa-a-ar-ša*) in two respects: the long vowel of the root is absent and there is additional syllable in the beginning. To understand this difference one has to reconstruct the Indo-European proto-forms. It is supposed that the root of the word had an initial laryngeal reflected in Old Indian *avatás* "spring", Latvian *avuõts* "spring" < \**H(e/o)w-nt-os* (making possible a reconstruction of a heteroclitc paradigm with the \**r/-nt-* alternation), Lehmann 1986:380 with

references. This laryngeal initial is confirmed by the ancient stems derived from the root. Thus for a cognate Hittite word *warša-* "dew" as well as for the related Greek *ἄερος* "dew", dialectal *ἄερος*, *ἄερος* (Chantraine 1990:375), Old Indian *varṣā-* "rain", one may suggest as a protoform either *\*h<sub>2</sub>wor-s-o-* (Eichner 1980:129; 1988:140; Melchert 1994:49) or *\*h<sub>2</sub>werse-* (Nussbaum 1986:125, 127; Bader 1992:399, n. 55). Assuming the root *\*Hew-*, two different stems with the suffixes ending in *-r* reflected in Luwian can be reconstructed:

1. Binominal structure of the first type (state I) according to Benveniste (Benveniste 1935; Gamkrelidze, Ivanov 1995:1, 194 ff.). The root which has the *\*e* grade is accentuated; the second non-accentuated morph has a syllabic resonant:

(I) *\*Héw-ṛ* > Luwian *uwar-* (for the sound laws *\*eu > u*, syllabic *\*ṛ > ar* cf. Melchert 1994; the resulting group *\*u + a- > uw-a* was changed after the Sievers-Edgerton rule).

This type is reflected in several other Luwian heteroclitic neuter nouns in *-ar*. In the case of *a-aš-ḫar-ša* [*ašḫar-ša*] (KUB XXXV 109 Rs.III 13; second half of the XVth century B.C.: Starke 1985:259, 266) the Luwian form corresponds to the Hittite *ešḫar* with the same archaic stress on the *e* (Hart 1980; Ivanov 1982; cf. Melchert 1994:235, 243, 263; Starke 1990:558). One may safely reconstruct the Indo-European initial stress on the root vowel *e* if not also its length possibly reflected in Greek: *ἔαρ*, *ἤαρ*, Old Indian *ásṛ-k* (Benveniste 1935:8, 26; Chantraine 1990:308).

The Luwian *ú-tar-ša* "word, spell" belongs to the same type. This spelling with the sign *ú* typical of the Luwian orthography (Rosenkranz 1952:33; cf. Otten 1953:97) is repeated 5 times: the above quoted old copy of the ritual of Puriyanni KUB XXXV 54 Vs.II 13', 38'; III 38', Starke 1985:66, 67, 69; a later copy of the same ritual dated by the end of the XIVth century KUB XXXV 55 9', Starke 1985:71; a ritual of the child-birth dated by the XIIIth century KUB 88 Rs. III, Starke 1985:227. In the last case another copy of the same ritual dated by the end of the XVth century has the parallel spelling *ú-ta-a[r-ša]* (KUB XXXV 89 2', Starke 1985:228, note 79). This variant is important for a comparison to the writing *ú-wa-ar-ša* since it shows that at least for the beginning of the New Hittite period the spelling by the combination of the signs Consonant + Vowel and Vowel + Consonant (CV + VC) was a normal way to render the unstressed short syllable (the second one in the stem of this morphonemic type).

Although the etymology of the Luwian *utar-* = Hittite *uddar* has remained controversial, still it can be assumed that it can be traced back either to *\*h<sub>2</sub>wodh<sub>2</sub>-* (Eichner 1980:129, n. 41; 1988:141; the comparison to Greek *αὖδῃ* was suggested already by Hrozný) or to *\*éutṛ-* (Melchert 1994:50, 126, 156, 242, 265) with the stress on the initial syllable having the full (*e* or *o*) grade of the root vowel whereas the second (suffixed) syllable has the syllabic resonant due to the zero or reduced grade.

The evidence of these forms make it plausible that this is the normal type of an old heteroclitic noun.

2. The structure of the second type with the zero grade of the unstressed root syllable and the stress on the second one having a long grade of the vowel *\*o* :

(II) *\*Hw-o:r > wa:r*

Formally this type coincides with the Hittite collective in *-ar* ("thème II à allongement radical avec *-r*", Benveniste 1935:181). The structure of the Hittite forms and their possible Indo-European origin were aptly formulated already 80 years ago by Hrozný. His remarks show the depth of his linguistic genius and deserve to be cited at length. Referring to the studies of Johannes Schmidt (1889) and Brugmann (1906-9) and quoting from the latter particularly the Avestan (*vi:spa:*) *aya:rə*: "(all the) days" (Bartholomae 1979:157; cf. Benveniste 1935, Nussbaum 1986:127, 129), Hrozný wrote: "Es ist unsicher, welche Ablautstufe in dem *-ar* des Sg. der in diesem Abschnitt behandelten heth. Wörter (*wa-a-tar...*, *ut-tar...* usw.) vorliegt (*\*-r?* *\*-er?*); in dem *-ar* von *ú-i-da-a-ar* usw. wird dagegen vermutlich ein urindogerm. *\*-o:r* zu erblicken sein (vgl. Brugmann ...578). Für das letztere ist einerseits auf das *-o:r* des Griech. Nom.-Akk.Sg. *ὕδωρ*, andererseits auf das *-a:r* des avest. Nom.-Akk. Pl. *aya:rə* hinzuweisen (vgl. l.c.). Nach J.Schmidt... hatten die Formen mit dehnstufiger Schlussilbe ursprünglich Kollektivbedeutung, so dass sie sowohl als Plural, als auch als Sing. verwendet werden konnten. Der Wechsel zwischen *wa-a-* (anscheinend -vielleicht bloss sekundäre?- Dehnstufe) und *ú-e/i-* (Reduktions- oder Vollstufe?) in der ersten Silbe des Wortes ist ebenfalls durch den indogerm. Ablaut zu erklären" (Hrozný 1916:65). Following this way of reasoning Hrozný assigned the forms like *ú-i-da-a-ar* (collective of "water") and *ud-da-a-ar* (collective of "word") both to singular and plural in the paradigms that he suggested (ib., 63-64, 67). The correctness of this view has been proved recently by the study of the writing devices like the relatively often used *INU-TIM ú-i-da-a-ar* (literally "one portion of water"), for instance KBo XXIII27+ Vs. III [21'], Rs. III21; see on this way to render collective forms Neu 1992:202-203, 211, n.24; XIV<sup>7</sup> TA.PAL *še-he-el-li-ya ú-i-da-a-ar* (KBo XXIV 45 Vs.32) "fourteen portions of the pure water" (ib.:206). On the base of such forms attempts have been made to reconstruct the Proto-Indo-European category of collective nouns opposed to the non-collective neutral stems, reviving the same ideas of Schmidt that had influenced Hrozný (Eichner 1985; Nussbaum 1986:118-130; Neu 1992; cf. already Tronsky 1946; 1967:66-69). The survival of the paradigmatic oppositions of the type of Hit. *watar* ~ *wida:r*, Luw. *uwar-* ~ *wa:r-* seems to belong to the common archaic features of the Hittite nominal system and the Luwian one. A similar opposition can be reconstructed for Proto-Tokharian where an old neutral proterokinetic noun *\*péHur > paur > Tokharian A por* "fire" (Þórhallsdóttir 1988:200) is related to the old collective *\*pHwo:r > puwo:r > B puwar* ( Hilmarsson 1986:207; 1989:21, 113, 135; Klingenschmitt 1994:400-401, n. 151; Schindler 1967:242-

244; 1975:10) in the way similar to the difference between the usual Hittite *pa-aḫ-ḫur* (starting with the Old period; see on the Middle Hittite plene-writing in KUB XVII 10 III 22 : Melchert 1994:147), Luvian *pa-aḫ-ḫu-u-ur* (< Proto-Anatolian *\*páHwṛ* according to Melchert 1994:55, 98,122) and the rare Hittite collective form *pa-aḫ-ḫu-wa-ar* (KUB VII 60 II 11). Although some traces of the same paradigm with the opposition of the proterokinetic non-collective neuter and the hystero-kinetic neuter collective (cf. Oettinger 1993) still have been preserved in such isolated cases as Greek double forms τέκμαρ “mark” /Homeric τέκμων “goal, end, token” (Benveniste 1935:20, 116, 121; Chantraine 1990:1099-1100; Nussbaum 1986:119-120 with a reference to Shindler’s talk at the Yale Linguistic club), in the other Indo-European dialects the whole system have been transformed (possibly due to the change of the gender system).

If in Luvian these oppositions have retained the same semantic value as in Hittite, one may search for their trace in the use of *wa:r-ša*. This form may refer to the portion of the water that has been poured into the clay bowl in the beginning of the ritual of Puriyanni. In that case one may translate the first Luvian sentence of the ritual quoted above as “This is the portion of water taken from the river.” In some cases the Vedic use of the cognate form *va:r* seems close to this hypothetical meaning. The Old Indian word is but rarely used in *Ṛg-Veda* (Grassmann 1873:1260; on some of the other ritual terms for water and examples of their use cf. Jamison 1996:127-149). If the purely anagrammatical play of the phonemes (as in IV,19,4: *vá:r na vá:tas* “as the wind on the water”, Elizarenkova 1989:382) and metaphorical images (*vá:r na pathá: ráthyeva sva:ni:t* “as the water produces noise by the road ...”, II, 4,6; also IX, 112,4; X, 145, 6 cf. ib., 241; Elizarenkova 1972:141) are not taken into consideration, in some other cases clearly a portion of water taken from some larger space has been meant: *ṣarásya cid a:rcatásya:vátá:d a ni:cá:d uccá: cakrathuḥ pá:tave vá:h* “You have got the drinkable water from the well -from the bottom up- for Šara, the son of Richatka” I, 116, 22, cf. Elizarenkova 1989:143. The image of the heavenly bucket (on the concept see Kuiper 1972; 1983) is meant in the hymn to Indra where after mentioning another term for water the word *va:r* is introduced as a designation of a basin (pool) to which the streams go (VIII,98, 6-7; Elizarenkova 1995:439). In the famous story of a maiden Apa:la: (for an interpretation of the ritualistic side of it see: Jamison 1991:149-172; 1996:240) one of the reasons for her to go down to the water (*va:r*, VIII, 91, 1) was to fetch some portion of it (Schmidt 1987:2-3,11); the Vedic opposition of the words for water and stream in this text coincide with a similar one in the Luvian text of the ritual of Puriyanni (Ved. *áp-i* = Luw. *ḫap(i)-*, Ved. *va:r* = Luw. *wa:r-*; the terms are opposed by the archaic feature active- non-active reflected in their gender).

It seems possible to continue the search for the traces of obsolete collective and non-collective heteroclitic nouns preserved in Hittite, Luvian and other archaic Indo-European languages. Some questions put by Hrozný in the above



mentioned study has still remained unanswered. Thus one can not find established the full grade *\*e* in the first syllable of the Hittite collective form *wida:r* (Shindler 1975:4; Eichner 1985:165; Nussbaum 1986:127-129; Melchert 1994:106). A typologically similar Luwian *uwar-* as well as other collective forms quoted above present a zero grade in the first syllable. The view of *i* in *widar* as a trace of a reduced vowel (Benveniste 1935:26) may be supported by the Old Hittite evidence on the same function of *i* in the paradigm of *ešhar/išha-* (Rosenkranz 1978:35, 50; cf. also a reflex of the same vowel in the initial syllable and of the long grade in the second part of the form in the ancient Homeric Greek borrowing from a dialect of a Hittite type: *ἰχῶρ* "the blood of the gods", Gamkrelidze, Ivanov 1995:1, 798 with references; the word seems to go back to a collective *\*sHo:r*, Tokharian A *ysa:r*, B *ysar*, cf. Klingenschmitt 1994:396, n. 140; Nussbaum 1986:123). The old character of the vowel *i* in the collective forms might have been confirmed by the Palaic *ša-a-ú-i-da-a-ar* "horn, plenty" where the supposition of the influence of the Hit. *ú-i-da-a-ar* would have led to the hypothesis on the Proto-Anatolian age of the latter (Oettinger 1989:202), although in an old *\*-tro-* derivative from a "disyllabic base" (Hit. *šuwa-*, ib.) one might expect an *-i-* of the type similar to the Old Indian *ar-í-tra-* (cf. Burrow 1979:84). A less ambivalent collective form is represented in Luwian *huitar* "wild animals". If it can be derived from *\*Hwed-* (Melchert 1994:262, 273), it may support a hypothesis on the archaic character of the reduced grade leading to *i*.

As Luwian seems to reflect some ancient structural features of the Indo-European opposition of the collective and non-collective neutral forms, the data of this language might help in revising the reconstruction of this part of the ancient nominal system.

An interesting problem of the Indo-European linguistic geography concerns the distribution of the neutral nouns derived from the root *\*h<sub>2</sub>ew-* in the meaning "water". Only in Luwian, Tokharian and Old Indian have the derivatives with the suffix *\*-o/o:r* been found in this function. The form of the stems in Luwian and Vedic Sanskrit is similar (old collective *\*h<sub>2</sub>w-o:r*), in Tokharian the stem *\*h<sub>2</sub>w<sub>e</sub>r- > A wär, B war* arose due to the influence of the secondary suffixes (van Windekens 1976:13, 14, 44 and 558; see there also on the impossibility of supposing the loss of *\*-d-* in the word). In most other Indo-European languages as well as in Old Indian, one finds the nouns derived from the same root and the suffix *\*-e/o/0d-* (known also in the verbs with a nasal infix, i.e. suffix) after which secondary heteroclitic elements *\*-e/o/0r~n* follow: *\*h<sub>2</sub>eu-d-* (+ *\*r*), *\*h<sub>2</sub>w-ed-*, *\*h<sub>2</sub>u-d-e/o(:)r-* etc. (Benveniste 1935:26, 159, 183; cf. Strunk 1972:175; Bader 1992:388-389; Gamkrelidze, Ivanov 1995:216). The distribution makes it possible that the Luwian and Old Indian forms corresponding to each other represent a more archaic structure.

## References

Bader, F.

- 1992 "Boire de l'eau." *Perspectives on Indo-European Language, Culture and Religion. Studies in Honor of Edgar C. Polomé*. Journal of Indo-European Studies. Monograph No. 9, McLean, Virginia: Institute for the Study of Man, 380-405.

Bajun, L.S.

- 1978 "Ob odnom slučae morfologičeskoj interferencii (luvijskij padež na -ša)." *Issledovanija po fonologii i grammatike vostočnyx jazykov*. Moskva.

Bartholomae, Chr.

- 1979 *Altiranisches Wörterbuch zusammen mit Nacharbeiten und Vorarbeiten*. Berlin, New York: Walter de Gruyter.

Benveniste, E.

- 1935 *Origines de la formation des noms en indo-européen*. Paris: Adrienne Maisonneuve (3rd printing, 1962).

Brugmann, K.

- 1906-9 *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. 2 Band 1, 2/3 Teil. Strassburg.

Burde, C.

- 1974 *Hethitische Medizinische Texte. Studien zu den Bogazköy-Texten*. Heft 19. Wiesbaden : Otto Harrassowitz.

Burrow, T.

- 1979 *The Problem of Schwa in Sanskrit*. Oxford: Clarendon Press.

Carruba, O.

- 1982 "Der Casus auf -sa des Luwischen." *Investigationes philologicae et comparativae. Gedenkschrift für Heinz Kronasser*. Ed. E. Neu. Wiesbaden : Otto Harrassowitz: 1-15.
- 1982a "Beiträge zum Luwischen." *Serta Indogermanica. Festschrift für G. Neumann zum 60. Geburtstag*. Ed. J. Tischler. Innsbruck, 35-51.
- 1992 "Le notazioni dell'agente animato nelle lingue anatoliche (e l'agente animato nelle lingue anatoliche e l'ergativo)." In: Carruba (ed.) 1992, 61-98.
- 1992 (ed.) *Per una grammatica ittita. Towards a Hittite Grammar*. Studia Mediterranea 7. Pavia: Gianni Iuculano Editore.

Chantraine, P.

- 1990 *Dictionnaire étymologique de la langue grecque*. Paris: Éditions

Klincksieck.

Dixon, R.M.W.

- 1994 *Ergativity*. Cambridge Studies in Linguistics, 69. Cambridge: Cambridge University Press.

Eichner, H.

- 1980 "Phonetik und Lautgesetze des Hethitischen." *Lautgeschichte und Etymologie*. Akten der VI. Fachtagung der indogermanischen Gesellschaft. Ed. M. Mayrhofer a.o., Wiesbaden: Reichert, 120-165.  
 1985 "Das Problem des Ansatzes eines urindogermanischen Numerus "Kollektiv" ("Komprehensiv")." *Grammatische Kategorien*. Akten der VII. Fachtagung der Indogermanischen Gesellschaft, ed. B. Schlerath. Wiesbaden, Reichert, 134-169.  
 1988 "Anatolisch und Trilyngalismus." *Die Laryngalthorie und die Rekonstruktion des indogermanischen Laut- und Formensystems*. Ed. A. Bammesberger. Heidelberg: Carl Winter-Universitätsverlag.

Elizarenkova T. Ja. (ed.)

- 1972 *Rigveda. Izbrannye gimny*. Moscow: Nauka.  
 1989, 1995 *Rigveda. Mandaly I-IV. Mandaly V-VIII*. Moscow: Nauka.

Gamkrelidze, T., Ivanov, V.

- 1995 *Indo-European and the Indo-Europeans* (Trends in Linguistics. Studies and Monographs 80, ed W. Winter). I-II. Berlin, New York: Mouton de Gruyter.

Garrett, A.

- 1990 "The origin of NP split ergativity." *Language*, vol. 66, 261-296.

Grassmann, H.

- 1873 *Wörterbuch zum Rig-Veda*. Leipzig: F. A. Brockhaus.

Güterbock, H.G. and Hoffner, H. (ed.)

- 1980 *The Hittite Dictionary of the Oriental Institute of the University of Chicago*, vol. 3, fasc. 1. Chicago : The Oriental Institute of the University of Chicago.

Hart, G.

- 1980 "Some observations on plene-writing in Hittite." *Bulletin of the School of Oriental and African Studies*, 43, 1-17.

Hilmarsson, J.

- 1986 *Studies in Tocharian Phonology, Morphology and Etymology*. Reykjavík.

- 1989 *The Dual Forms of Nouns and Pronouns in Tocharian*. Tocharian and Indo-European Studies Supplementary Series. Vol. 1. Reykjavík.
- van den Hout, T. P. J.  
 1984 "Einige luwische Neutra auf -ša/-za in überwiegend jung-hethitischen Texten." *KZ* 97, 60-80.
- Hrozný, F.  
 1916 *Die Sprache der Hethiter, ihr Bau und ihre Zugehörigkeit zum indogermanischen Sprachstamm. Ein Entzifferungsversuch*. Boghazköi-Studien, ed. O. Weber, 1 Heft. Leipzig: J.C. Hinrichs'sche Buchhandlung
- Ivanov, V. V.  
 1981 *Slavjanskij, baltijskij i rannebalkanskij glagol. Indoevropskie istoki*. Moscow: Nauka.  
 1982 "Novyj istočnik dlja ustanovlenija indoevropskix akcentuacionnyx paradigm." *Balto-slavjanskije issledovanija 1981*. Moscow: Nauka.  
 1996 "Iz zametok o praslavjanskix i indoevropskix čislitel'nyx." *Rusistika. Slavistika. Indoevropistika. Sbornik k 60-letiju A. A. Zaliznjaka*. Moscow: Indrik, 704-727.
- Jamison, S. W.  
 1991 *The ravenous hyenas and the wounded Sun: Myth and Ritual in ancient India*. Ithaca and London.  
 1996 *Sacrificed Wife and Sacrificer's Wife. Women, Ritual and Hospitality in Ancient India*. New York-Oxford: Oxford University Press.
- Klingenschmitt, G.  
 1994 "Das Tocharische in indogermanistischer Sicht." *Tocharisch. Akten der Fachtagung der Indogermanischen Gesellschaft*. Berlin, September 1990. Ed. B. Schlerath. Reykjavík: Málvísindastofnun Háskóla Íslands, 310-412.
- Kuiper, F. B. J.  
 1972 "The heavenly Bucket." *India Maior. Congratulatory volume presented to J. Gonda*. Leiden, 1972, 144-156.  
 1983 *Ancient Indian Cosmology*. Ed. J. Irwin. Delhi.
- Laroche, E.  
 1959 *Dictionnaire de la langue louvite* (Bibliothèque archéologique et historique de l'Institut Français d'archéologie d'Istanbul, VI). Paris: Librairie Adrien-Maisonneuve.

Lehmann, W.P.

1986 *A Gothic Etymological Dictionary*. Leiden: E.J. Brill.

Melchert, H. C.

1988 "Luvian Lexical notes." *Historische Sprachforschung*, 101, 211-243.

1993 *Cuneiform Luvian Lexicon*. Chapel Hill.

1994 *Anatolian Historical Phonology*. Leiden Studies in Indo-European 3, Amsterdam-Atlanta, GA: Rodopi.

Meriggi, P.

1957 "Zum Luwischen." *Wiener Zeitschrift für die Kunde des Morgenlandes*, Bd. 53, Heft 3-4, 193-226.

1980 "Schizzo grammaticale dell' anatolico." *Atti della Accademia Nazionale dei Lincei, Anno CCCLXXVII. Memorie. Classe di scienze morali, storiche e filologiche*, serie VIII, vol. XXIV, fasc. 3.

Neu, E.

1992 "Zum Kollektivum im Hethitischen." In: Carruba (ed.) 1992, 197-212.

Nussbaum, A. J.

1986 *Head and Horn in Indo-European*. Studies in Indo-European Language and Culture, New Series. 2. Berlin, New York: Walter de Gruyter.

Oettinger, N.

1989 "šawitra "Horn", eine hethitische \*-tro- Bildung." *Hethitisch und Indogermanisch*. Ed. E. Neu, W. Meid, 197-203.

1993 "Der Akzent des indogermanischen Kollektivums in Lichte des Hethitischen." *Münchener Studien zur Sprachwissenschaft*, 54: 207-14.

Otten, H.

1953 *Zur grammatikalischen und lexikalischen Bestimmung des Luwischen*. Deutsche Akademie der Wissenschaften zu Berlin. Institut für Orientalforschung. Veröffentlichung Nr. 19. Berlin: Akademieverlag.

Pokorny, J.

1959 *Indogermanisches etymologisches Wörterbuch*. Bern, Munich: Francke.

Renou, L.

1958 *Études védiques et paninéennes*. T. IV. Paris: Éd. E. de Boccard.

Rosenkranz, B.

1952 *Beiträge zur Erforschung des Luwischen*. Wiesbaden: Otto Harrassowitz.

1978 *Vergleichende Untersuchungen der altanatolischen Sprachen*. Trends in Linguistics. State-of-the-Art Reports 8. Ed. W. Winter. The Hague, Paris-New York: The Mouton Publishers.

Schindler, J.

- 1967 "Tocharische Miszellen." *Indogermanische Forschungen*, 72 Bd., 3H.:239- 249.  
 1975 "L'apophonie des thèmes indo-européens en -r/n." *Bulletin de la Société linguistique de Paris*. 70,1-10.

Schmidt, H.-P.

- 1987 *Some women's rites and rights in the Veda*. Post-graduate and research Department Series N 29. Professor P.D. Gune Memorial Lectures. Poona, India: Bhandarkar Oriental Research Institute.

Schmidt, J.

- 1889 *Die Pluralbildungen der indogermanischen Neutra*. Weimar: Böhlau.

Shevoroshkin, V.

- 1979 "On the Hittite-Luwian numerals." *Journal of Indo-European Studies*, vol.7, 177-198.

Starke, F.

- 1985 *Die Keilschrift-luwischen Texte in Umschrift*. Studien zu den Bogazköy-Texten 31. Wiesbaden: Harrassowitz.  
 1990 *Untersuchung zur Stammbildung des keilschrift-luwischen Nomens*. Studien zu den Bogazköy-Texten 31. Wiesbaden: Harrassowitz.

Strunk, K.

- 1972 "Neue Gesichtspunkte zu Genesis und Struktur von Nasalpräsentien nach Art des altindischen 7. Klasse." *Folia Linguistica*, 4, 1/2, 175-178.

Toporov, V.N.

- 1975 *Prusskij jazyk. Slovar'. A-D*. Moscow: Nauka.

Tronsky, I.M.

- 1946 "K semantike množestvennogo čisla v grečeskom i latinskom jazykax." *Učenyje zapiski Leningradskogo Universiteta*, vyp. 10:54-72.  
 1967 *Obščeeindoevropskoe jazykovoe sostojanie*. Leningrad: Nauka, Leningradskoe otdelenie.

Turner, R.L.

- 1989 *A Comparative Dictionary of the Indo-Aryan Languages*. London- New York- Toronto: Oxford University Press.

þórhallsdóttir, G.

- 1988 "Tocharian Contraction across -w-." *Tocharian and Indo-European Studies*. Vol. 2:184-209.

Watkins, C.

- 1987 "Two Anatolian forms: Palaic *aškummauwa*, Cuneiform Luwian *wa-a-ar-ša*." *Festschrift for Henry Hoenigswald on the Occasion of his 70th Birthday*, ed. by G. Cardona and N. Zide, Tübingen: Narr, 399-404 (=Watkins 1994, 309-314).
- 1994 *Selected Writings*. Innsbrucker Beiträge zur Sprachwissenschaft. 80. Innsbruck.
- 1995 *How to kill a Dragon: Aspects of Indo-European Poetics*. New York, Oxford: Oxford University Press.

van Windekens, A.J.

- 1976 *Le tokharien confronté avec les autres langues indo-européennes*. Travaux publiés par le Centre International de Dialectologie générale de l'Université catholique néerlandaise de Louvain. Vol. I. *La phonétique et le vocabulaire*. Louvain : Centre International de dialectologie générale.

# Macrorelationships and Microrelationships and their Relationship\*

Brian D. Joseph  
*The Ohio State University*

Macrorelationships, i.e. relationships between languages and language families at a time-depth beyond what is normally deemed (easily) reachable with the Comparative Method,<sup>1</sup> have loomed large in both the historical linguistic literature and the popular literature on linguistics in recent years, due largely to the interest in the topic provoked by Joseph Greenberg's 1987 book *Language in the Americas* and other more recent pieces in a similar vein.<sup>2</sup>

These macrorelationships are often speculative though they generally have a ring of truth — or at least, plausibility — about them. What is especially tricky about claims of such "long-distance" relationships is that they are hard to prove. Thus, much of the debate about these claims has concerned the meaning of "proof" in this domain, focusing on how much evidence is enough, what type of evidence is probative, and what the nature is of the methodology that leads to or supports the claimed relationships.

The focus on the evidence and how to evaluate it means that an interesting and revealing comparison can be made between the methods for judging macrorelationships and the methods for determining what can be referred to as microrelationships, i.e. subgroupings within well-established or well-recognized linguistic groups. Especially interesting in this regard are those cases in which the degree and/or nature of the microrelationship is unclear, whether because of a general lack of data, an absence of just those crucial data points needed to clinch the argument one way or the other, or some similar obscuring factor.

---

\*A version of this paper was presented at the First Workshop on Comparative Linguistics, held in November 1992 at the University of Michigan, in which Vitaly Shevoroshkin was an active and important participant. I have benefitted from comments by Sasha Vovin, Sheila Embleton, and Eric Hamp after that presentation. Much of the material contained herein is based on joint work I have done with Rex E. Wallace of the University of Massachusetts (see the references for relevant bibliography), though he is not responsible for the uses I have put it to here.

<sup>1</sup>I say "normally deemed" here to reflect a general belief (see, for instance, Nichols 1992: 5-6, 184, Ruhlen 1994: 14 for some discussion and references) that the utility of the comparative method diminishes when comparisons are at a time-depth greater than some rather large number (around 10,000 years is the figure often mentioned). I take no stand on this claim, but note that it may well simply be a practical constraint based on the difficulty of finding reliable comparanda at such a time-depth rather than an absolute constraint inherent in the method itself.

<sup>2</sup>See, for instance, Campbell 1988, Greenberg 1989, Greenberg & Ruhlen 1992, Matisoff 1990, Ross 1991, and Ruhlen 1994, as well as work by the honoree of this volume, e.g. Shevoroshkin 1989a, 1989b, 1989c, 1990a, 1990b, 1991.



There are several reasons for exploring this comparison between microrelationships and macrorelationships and the methods they require. First of all, both types of relationships are exercises in the classification of languages: subgrouping involves family-internal classification, whereas long-distance relationships involve connections among families. The similarity is evident when one considers that if something like "Proto-World" is correct, and all the languages of the world are related to one another,<sup>3</sup> then all relationships would really turn out to be a type of subgrouping, for the issue would not be whether two languages are related at all,<sup>4</sup> but rather how closely they are related, i.e. a question of subgrouping.<sup>5</sup>

Second, without getting into the thorny issue of whether reconstruction is necessary to prove a claim of relationship, it is clear that positing a linguistic relationship is intimately connected to being able to reconstruct linguistic features of common ancestor, the proto-language, for the languages in question. Macrorelationships ultimately lead one to attempt reconstruction,<sup>6</sup> but in doing reconstruction at the microrelationship level, i.e. within "lower-order" language families, it is essential to get the subgrouping right. In fact, successful reconstruction depends on the determination of the subgrouping relationships within a family, for it is not possible to judge adequately how widespread an innovation is without a sense of what the finer degrees of relatedness are among members of the family. For example, the labial correspondences within Indo-European, when arranged as in (1), present a primarily even mix of fricatives and stops, thus getting in the way of a clear decision as to what to reconstruct:

1. English **f** = Greek **p** = Irish **Ø** = German **f** = Russian **p** = Armenian  
**Ø/h** = Gothic **f** = Latin **p** = Avestan **f** (but only before a

<sup>3</sup>More accurately, as noted in Hock & Joseph (1996:488), the issue of Proto-World is really a matter of whether all oral languages are related, for discussions of Proto-World have generally ignored the many signed languages that have developed within the history of the world.

<sup>4</sup>It is useful here to note that though it is often said that it cannot be shown that two languages are not related, in fact there are pairs of languages that simply cannot be related, specifically Esperanto (especially among those who (reportedly) use it as their first language) and a signed language such as American Sign Language.

<sup>5</sup>As Ruhlen (1994: 272) puts it: "it no longer makes sense to ask if two languages (or language families) are related. *Everything* is related, and the question to be investigated within or among different families is the *degree* of their relationship, not the fact of it".

<sup>6</sup>For example, the extent of reconstruction attempted for Nostratic is a case in point. It has never been enough to simply claim that Indo-European, Uralic, Kartvelian, etc. are related; rather, serious discussion of Nostratic has involved reconstruction of the proto-language as well. See for instance recent works such as Manaster Ramer, Michalove, Baertsch, & Adams 1997 and Bomhard & Kerns 1994, as well as papers in Shevoroshkin 1989b, 1989c, 1990b, 1991.

consonant) = Sanskrit **p** = Albanian **p** = Old Norse **f** = Hittite **p** = Tocharian **p**

However, once subgrouping, based on independent criteria, is taken into account, as in (2), stop reflexes predominate and thus the reconstruction of \***f** becomes somewhat less plausible<sup>7</sup>:

2. Germanic \***f** (= English **f**, German **f**, Gothic **f**, Old Norse **f** (etc.)) = Indo-Iranian \***p** (= Avestan **f** /    **C** = Sanskrit **p**) = Greek **p** = Irish **Ø** = Russian **p** = Armenian **Ø/h** = Latin **p** = Albanian **p** = Hittite **p** = Tocharian **p**

Third, given the recent attention to macrorelationships, any new perspective offered by microrelationships on the subject ought to be important, especially since the methodologies in both pursuits are quite parallel. In particular, in doing subgrouping, especially in the unclear cases, the evidence is often quite slim, and open to conflicting interpretations; also, given such evidence, the basic principle of classification, i.e. to pay attention to shared innovations, and especially to shared particularities of development, can often be obscured by mere shared similarities, i.e. by similarities of form that do not necessarily point to a common parentage for the languages involved.

To be sure, there is still a lot of work to be done on microrelationships. One need only consider the fact that within a relatively well-studied language family like Indo-European, numerous controversies regarding subgrouping remain to be worked out, for instance, whether there was an Italo-Celtic subgroup,<sup>8</sup> what the relationship was between Greek and ancient Macedonian, where Old Prussian fits in within Baltic and more generally within Balto-Slavic, if there even is a Balto-Slavic subgroup, and so on.

The pitfalls of working on microrelationships are illustrated here by the examination of one problematic case in some depth, which then allows for some explicit parallels with the enterprise of hunting for macrorelationships. The case in point is the relationship between Latin and Faliscan, two languages spoken in ancient Italy, though reference to their relationship with Oscan and Umbrian, two other languages of ancient Italy, is also relevant.

Mention of all these considerations should not be taken as support for a view that one cannot proceed with any long-distance relationships until all the

<sup>7</sup>This is not to suggest that decisions about what to reconstruct are simply a numbers game, with the majority reflex chosen as the proto-language element; rather, a number of criteria need to be taken into consideration. Still, it is safe to assume that most practicing historical linguists would be more inclined to reconstruct a stop when confronted with the correspondences in (2) than with those in (1); thus, there is some safety in numbers in reconstruction, but some pitfalls as well.

<sup>8</sup>Or even an Italic branch, as opposed to separate Latin and Oscan-Umbrian branches stemming directly from Proto-Indo-European, a position evaluated below.

details of closer-range relationships are cleared up. On the contrary, both pursuits should proceed for they can feed into one another, but that there are enough parallels in the methodologies — in fact, as noted above, virtually the exact same methodologies are needed — to allow for progress to be made by learning from both enterprises.

To turn first to the relationship between Latin and Oscan-Umbrian, two main claims can be recognized in the literature: one that has Latin and Oscan-Umbrian subgrouped together as an Italic branch of Indo-European, as indicated in (3a), and one that treats Latin and Oscan-Umbrian as separate branches of Indo-European, each on a par with Greek or Indo-Iranian, as indicated in (3b):<sup>9</sup>

3. a.
- ```

      Proto-Indo-European
      /      \
    Greek   Italic .....
            /  \
           /    \
        Latin Oscan-Umbrian
  
```
- b.
- ```

      Proto-Indo-European
      /      |      \
    /      |      \
  Greek   Latin Oscan-Umbrian ...
  
```

Although there are several potential shared features worthy of investigation as possible evidence to decide between (3a) and (3b),<sup>10</sup> one striking similarity between Latin and Oscan-Umbrian can be explored here: the form of the first person singular (1SG) present indicative of the verb 'be', *sum* in Latin and *súm* (phonetically [som]) in Oscan. This similarity at first glance would suggest a reconstruction for Italic of \*som, which would represent a significant deviation, an apparent shared innovation, away from PIE \*(H<sub>1</sub>)esmi (as defined by the equation of Greek εἰμί = Sanskrit *asmi* = Gothic *am*, etc.).

One scholar, however, Bader 1976, has argued instead that the *sum/súm* similarity represents a shared retention between Latin and Oscan, and is thus insignificant for subgrouping. Interestingly, from the perspective of methods used in positing some macrorelationships,<sup>11</sup> this claim rests entirely on a false

<sup>9</sup>See Joseph & Wallace 1987 for discussion of these positions, with literature. Here and elsewhere in this paper, I use the traditional label "Oscan-Umbrian" instead of the now more usual "Sabellian" to allow for a greater point of contact with the previous work cited.

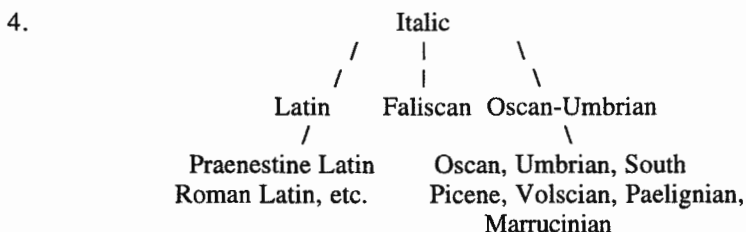
<sup>10</sup>For instance, the organization of the verbal system into four major conjugational classes seems like a significant shared innovation, and numerous others have been proposed, some of which are discussed in Joseph & Wallace 1987.

<sup>11</sup>For instance, Campbell (1988: 605-6) notes that several of the forms Greenberg 1987 cites in support of his Amerind hypothesis, whereby a good many of the

segmentation that Bader made: only if Tocharian B *nasam* 'I am' is analyzed as *na-sam*, with an "empty" preverb \*no- (as in the Irish imperfect), does the shared retention hypothesis gain some credibility. However, the Tocharian-internal evidence points to root \*nes- and the segmentation *nas-am*: Tocharian A 1SG *nesau*, 2SG *nest*, etc., Tocharian B 1SG *nasam*, 2SG *nast*, etc.

Still, there is one way in which Bader was right, in that the *sum/súm* similarity does not in and of itself represent a shared innovation. Rather, as Joseph & Wallace 1987 argue, the best account takes each form to be an independent outcome of forms that resulted from a few real shared innovations: enclisis of 'be', giving enclitic 1SG \*X-*esmi*, followed by loss of final \*-i in present tense verb forms, giving \*X-*esm*, and then by epenthesis *cum* rounding to give \*X-*esom*, all taking place in Common Italic, with the development to Latin *sum* and Oscan *súm* then being the result of similar but distinct processes within each language. In that case, *sum/súm* do point to a Latin-Oscan-Umbrian subgrouping, but not because they are so similar in form; a bit of digging shows that there is a significant shared innovation (actually a few) lurking behind them, but the obvious one is not significant.

With an Italic branch thus established (and see also footnote 10), the relationship of Latin to Faliscan can be considered. Several proposals for this relationship have been put forth. One is that of Beeler 1956, who treated Faliscan as equal sibling to all of Latinity, i.e. to the collection of various Latin dialects, including the Latin of Rome, of Praeneste, etc., and thus on an equal footing with Oscan-Umbrian, as modeled in (4):

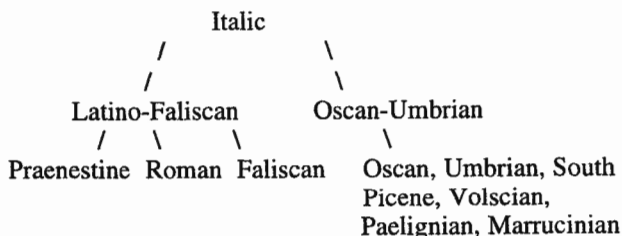


Another position, that of Beeler 1963, Campanile 1961, Eska 1987, Giacomelli 1979, Palmer 1954, Pisani 1962, and Pulgram 1978, treats Faliscan as a dialect of Latin, parallel to the Roman Latin, Praenestine Latin, etc., as illustrated in (5):

---

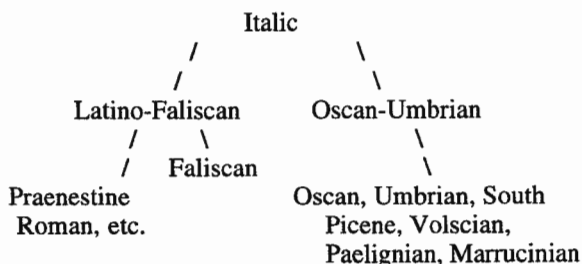
languages of the Americas belong to a single language family he calls "Amerind", in fact represent erroneous segmentations on Greenberg's part; see also Rankin (1992: 339).

5.



Finally, Faliscan has been considered, e.g. by Leumann 1977 and Sommer 1977, to be a sibling to all of Latinity within a Latino-Faliscan subgroup of Italic, modeled in (6):

6.



There is evidence that can decide among the three possibilities sketched in (4) through (6). Since this evidence is discussed thoroughly in Joseph & Wallace 1991, it is presented here in schematic form. In particular, a couple of innovations shared by Latin and Faliscan argue against (4) and thus in favor of a Latin-Faliscan subgroup; for instance, as indicated in (7a),<sup>12</sup> Faliscan and Latin have a future marker with an initial labial, as opposed to an *s*-marker in Oscan-Umbrian, the apparent inherited Italic norm, to judge from the occurrence of *s*-futures elsewhere, such as in Greek. Similarly, Latin and Faliscan show a \*-*d* suffix in the accusative singular of personal pronouns, where Oscan-Umbrian have *-om*, a marker paralleled elsewhere, e.g. in Sanskrit, as indicated in (7b):

7. a. *f/b*-future: Faliscan *carefo* LF 5, Latin *carebo* 'I will lack' (vs. Oscan-Umbrian *-s*-future; cf. Greek *s*-future)  
 b. \*-*d* in ACC SG of personal pronouns: Faliscan *med* LF 1, Latin *med* 'me' (vs. *-om* in Oscan-Umbrian, e.g. Umbrian *tiom* 'you'; cf. Sanskrit *mām* / *tvām* 'me / you')

<sup>12</sup>Sources of the forms cited in this and following displays are indicated by abbreviations: CIE = Corpus Inscriptionum Etruscarum; CIL = Corpus Inscriptionum Latinarum; LF = Giacomelli 1963; M = Marinetti 1985; TLE = Pallottino 1968<sup>2</sup>; Ve = Vetter 1953.

Interestingly, in terms of the quality of the data that can be used here and how frustrating fragmentary evidence can be if strict criteria are adhered to for relatedness and/or subgrouping, two tantalizing lexical parallels can be cited between Latin and Faliscan which prove to be unusable. In particular, the languages agree on the word for 'tomorrow', Faliscan *cra* (LF 5) and Latin *cras*, and on the word for 'today', Faliscan *foied* (LF 5) and Latin *hodie* 'today'. This latter form would be especially interesting if the Faliscan *-o-* were short,<sup>13</sup> but there is no way to tell, due to nature of Faliscan orthography. Even so, neither form can be used for determining an especially close relationship between Latin and Faliscan since the Oscan-Umbrian word for 'tomorrow' and 'today' are not known; thus, even though these forms are unparalleled elsewhere in Indo-European, it is not clear if they represent Latino-Faliscan innovations or Italic ones. In the absence of such information, these parallels remain tantalizing but inconclusive.

Among the evidence that has been invoked to support the position in (5), in which Faliscan is (just) a dialect of Latin, are the features cited in (8), each with an example from Faliscan and one from "dialectal" Latin (i.e. non-Roman Latin), contrasted with a Roman Latin form (simply labeled "Latin" here):

8. a. \*erC > irC: Faliscan *loifirtato* LF 25, [l]oifirta LF 73, Dialectal Latin *mircurios* CIL 1<sup>2</sup>, 564 [Praeneste] vs. Latin *libertas*
- b. monophthongization of diphthongs: Faliscan *efiles* LF 15, *pola* LF 74, Dialectal Latin *edus* [Varro, *LL* 5, 97]; *plotia* CIL 14, 3369 [Praeneste] vs. Latin *aediles*, *Paulla*, *haedus*, *Plautia*
- c. \*iV > eV: Faliscan *hileo* LF 97, *filea* LF 67, Dialectal Latin *fileai* CIL 1<sup>2</sup>, 561 [Praeneste] vs. Latin *filius*, *filia*
- d. loss of word-final consonants: Faliscan *mate* LF 121.1, *cupa* LF 121.1, Dialectal Latin *maio* CIL 1<sup>2</sup>, 76 [Praeneste]; *dedi* CIL 1<sup>2</sup>, 60 [Praeneste], *dede* CIL 1<sup>2</sup>, 47 [Tibur], [d]edero CIL 14, 2891 [Praeneste] vs. Latin *mater*, *cubat*, *maior*, *dedit*, *dederunt*
- e. *f* (vs. Latin *b/d*) in medial position from PIE aspirates: Faliscan *efiles* LF 15, *carefo* LF 5, Dialectal Latin *rufus* vs. Latin *aediles*, *carebo*, *ruber*

<sup>13</sup>The short *-o-* of Latin *hodie* preserves an archaic feature of Indo-European morphology, but in view of the variety of formations in words for 'today' in Indo-European languages (e.g. Greek *σήμερον/τήμερον* from \*ky-āmer-o-, Sanskrit *adya* from \*e-dye, etc.), it is likely that the compounding formation seen in the Italic words, if Faliscan has a short vowel and thus the forms are to be directly compared, is an innovation; that would make an agreement between Latin and Faliscan on this point potentially quite significant, dating to a time when their morphology still allowed the short vowel, though the Oscan-Umbrian forms would still be crucial to know.

- f. *f* > *h* in word-initial position (with hypercorrection of etymological *h* to *f*): Faliscan *hileo* LF 97 (*fe* LF 144), Dialectal Latin *horda* [Varro RR 2.5.6] (*faedus* Varro LL V, 97) vs. Latin *filius*, *hic*, *forda*
- g. Consonant-stem GEN in *-os* (> *-us*): Faliscan *lartos* LF 4a, *loifirtato* LF 25, Dialectal Latin *salutus* CIL 1<sup>2</sup>, 62 [Praeneste] vs. Latin *libertatis*, *salutis*
- h. *o*-stem GEN in *-osio*: Faliscan *kaisiosio* LF 4b, Dialectal Latin *popliosio ualesiosio* CIL 1<sup>2</sup> (4) 2832a [Satricum] vs. Latin *Caesi*, *Publi Valeri*

Significantly, however, all of these features in (8a) through (8h) are inadmissible as evidence bearing on subgrouping. They fail to pass muster against criteria for evaluating their utility in judging microrelationships: (i) a shared feature in and of itself is not probative if found in other languages, for if the languages are related, such a feature could be a common inheritance, and if they are not related (or even if they are) it could be the result of areal diffusion with the appropriate geographical distribution; (ii) a shared feature is significant generally only if it is a shared innovation, as noted already with regard to Latin and Oscan-Umbrian (cf. Hoenigswald 1960); (iii) even so, careful attention must be paid to the chronology of the features in question.

As it turns out, all of the features in (8) are problematic for one reason or another. The chronology of Faliscan features shows that some date to c. 300 B.C. or later, and thus are not old enough to be significant for determining relationship of Latin to Faliscan in Stammbaum terms, for the split of the two would necessarily predate the period of later similarity; these are listed in (9), along with the relevant data and their dates of attestation:

- 9. a. re (8b) Archaic Faliscan: *karai* LF 1 [c. 650 B.C.], *sociai* LF 3 [6th c. B.C.] Medio-Faliscan: *kaisiosio* LF 4b [5th c. B.C.]
- b. re (8c) Archaic Faliscan: *prauios* LF 1 [c. 650], *rufia* LF 3 [6th c. B.C.], *kalketia* LF 3 [6th c. B.C.]
- c. re (8d) Archaic Faliscan: *porded* LF 1 [c. 650], *ffiffliqod* LF 1 [c. 650], *fifiked* LF 11 [c. 500]
- d. re (8f) Archaic Faliscan: *far* LF 1 [c. 650], *ffiffliqod* LF 1 [c. 650], *huti[c]ilom* LF 1 [c. 650]

Thus these forms come after any period of presumed unity of Latin and Faliscan and therefore are not relevant for subgrouping, just as similarities among Greek, Albanian, Bulgarian, and the other modern languages of the Balkans, as members of the Balkan Sprachbund, are irrelevant for their place within the Indo-European family.

Others of these features reflect shared retentions, and thus as inheritances from (dialectal) Proto-Indo-European they are not significant for subgrouping. In particular, consonant-stem genitive singular forms in *\*-os* (see (8g)) are found in

Greek, e.g. κυνός 'dog/GEN', and in any case, Faliscan has consonant-stem genitives from \*-es, e.g. *f(e)licinate* (LF 73.2), just as Roman Latin does (cf. *libertat-is* in (8g)). Similarly, *o*-stem genitive singular forms in \*-osyo occur in Sanskrit, e.g. *devasya* 'god/GEN', in the Homeric Greek GEN.SG ending -οιο, and in the Armenian ending -oy.

In addition, the remaining features in (8), as well as some already shown to be irrelevant, are found all over ancient Italy, as indicated in (10), suggesting that they could be the result of areal diffusion or alternatively are relatively common developments; in either case, they would not be significant for subgrouping (the language, the source of the citation, and in some cases the place of attestation are noted):

10. a. re feature (8a): Oscan *amirikum* Ve 3, *mirikui* Ve 136
- b. re feature (8b): Umbrian *tota* VIa 29 < \*touta, Volscian *toticu* Ve 222 < \*toutikod, Marsian *pucle[s]* Ve 224 < \*putlois, Etruscan masculine praenomen *cnaive* TLE 14, Capua > *cneve* TLE 300, Volcii
- c. re feature (8c): Oscan *ionc* Ve 2 < \*eyom-ke, Marrucinian *iafc* Ve 218 < \*eyans-ke, Umbrian *tursiandu* < \*torseyantor
- d. re feature (8d): Umbrian *façia* IIa 17 < \*fakyad, Volscian *facia* Ve 222 < \*fakyad, Paelignian *dida* Ve 213 < \*didad, Marrucinian *pacrsi* Ve 218 < \*pakri sid
- e. re feature (8e): all except Roman Latin: Oscan *mefiaí* Ve 1, South Picene *mefiín* (M 1) (cf. Latin *medius*), Umbrian *alfu* Ib 29 (cf. Latin *albus*), Paelignian *loufir* Ve 209 (cf. Latin *liber*), Dialectal Latin *rufus* < \*H<sub>1</sub>reudhos (Latin *ruber* < \*H<sub>1</sub>rudhros)
- f. re feature (8f): Etruscan gentilicium *fuluna* (TLE 401, Volaterrae III-I) > *hulunias* (CIE 1900, Clusium III-I), cf. *vhulvena* (CIE 4952, Orvieto VI), and with hypercorrection *ferclite* (CIE 1487, Clusium III-I) for *herclite* (CIE 1486, Clusium III-I), from Greek Ἡρακλείδης.

A final problem with the features in (8) is that many of them in fact can be found in or attributed to Roman Latin, apparently reflecting a later transformation of original regional dialects into socially determined dialects within Rome itself, as Rome underwent extensive urbanization (see Joseph & Wallace 1992). As such, they are not really probative for determining Latin dialect groups definitively, unless one is able to successfully abstract away from the sociolects of Republican Rome, a difficult task to say the least. The relevant evidence of Roman attestations of these features is given in (11):

11. a. re feature (8b): monophthongization: *Pola* CIL 1<sup>2</sup>, 379 [Pisaurum, a Roman citizen colony], *Cesula* CIL 1<sup>2</sup>, 376 [Pisaurum]
- b. re feature (8d): loss of word-final consonants: *dedero* CIL 1<sup>2</sup>, 59



- c. re feature (8g): C-stem GEN in *-os* (> *-us*): *nominus* CIL 1<sup>2</sup>, 581 [Senatus Consultum de Bacchanalibus]

To work towards a solution to the question of how to determine the relationship of Latin and Faliscan, it would be desirable to focus on shared innovations that set Latin and Faliscan off from one another, especially innovations shared by Roman Latin and Dialectal Latin to exclusion of Faliscan, but even some innovations in Faliscan to exclusion of Roman and Dialectal Latin. The first type of situation would show that all varieties of Latin acted as a unity with respect to certain innovations, a unity that Faliscan did not participate in. The second type of situation is, on the face of it, less indicative of a Faliscan - Latin split, since in any of the models sketched in (4) through (6), Faliscan ultimately stands alone and thus can innovate to the exclusion of any other dialects or languages without those other speech communities having any particular unity amongst themselves. Still, under certain circumstances, Faliscan-only innovations can be significant, for instance when the other speech communities in question show a different innovation; in such a case, what is really involved, then, is an elaborated version of the first type considered.

Latin and Faliscan offer two possible examples of the first type, thus pointing to Faliscan being separate from all of Latinity. The first is the development of the PIE palatal voiced aspirate \*g'h, for it became *f* before *u* in all attested Latin but shows up as *h* in that position in Faliscan, and significantly, the Faliscan evidence comes from Archaic Faliscan of the 7th century before the Faliscan-internal change of *f* to *h* (see (8f) and (9f)), as shown by Latin *futis* 'water vessel' versus Faliscan *huti[c]ilom* 'vasette', both from PIE \*g'hu- 'pour' (whether PIE \*g'h went through Proto-Italic \*x or \*γ<sup>14</sup>). A second possibility is the innovative use all throughout Latin of *iacet* for 'lies', in place of inherited \*legh-, versus Faliscan *lece*t (LF 85), which is from \*legh-; the fact that this root is found in Latin in the noun *lectus* 'bed' (cf. Greek λέχος 'bed') is irrelevant, for there is no trace of a verb from \*legh- in Latin, and the replacement of the verb is the relevant innovation.

A few words of caution on these innovations are in order. For the first to be significant, it must be assumed that *futis* indeed is proper for all of Latinity and that had some Latin dialect had *hu-* for this word, some mention of it would have been made by some ancient grammarian (as is the case with some such forms

<sup>14</sup>See Wallace & Joseph 1993 for some discussion of the development of PIE \*g'h in Proto-Italic. I am assuming here that the change went through a Proto-Italic stage of a voiceless velar fricative, \*x, so that the Faliscan *h* reflects essentially no change from Proto-Italic, whereas the Latin *f* constitutes an innovation; if Faliscan *h* is judged to be an innovation (we do not really know what the exact phonetics of the Faliscan grapheme < H > were, after all, any more than we do for early Germanic *h* from Proto-Germanic \*x from PIE \*k), then this example is actually more like the second type discussed, where both groups show an innovative shift away from the starting point. Still, the import of the example for the ultimate point regarding the relationship of Faliscan and Latin is not affected.

known now only through such ancient indirect testimony). While this is not a difficult assumption to make, it nonetheless means that the interpretation is only as secure as this assumption. As for the second one, admittedly it does involve a lexical innovation and that weakens its import; since lexical items are so prone to being the material of borrowing, it cannot be ruled out that the innovative use of *iacet* is proper to just one dialect of Latin and that it was borrowed by others. If that were the case, *iacet* would not be a significant shared innovation in all of Latinity to the exclusion of Faliscan.

Still, overall, this evidence is highly suggestive of a significant separation between Latin and Faliscan, and further, these languages provide a reasonably solid example of the second type as well, strengthening the case even more.<sup>15</sup> In particular, Faliscan shows one innovation and all of Latin shows another, so that both deviate from a common Proto-Indo-European starting point. The development in question again involves PIE \*g'h, though this time in medial position. Whether through Proto-Italic \*x or \*γ, it develops into Latin *h*, as in *ueho* 'transport' < \*weg'h-, while in Faliscan, the outcome is *g* (spelled < q >, < c >, or < k >), as in *lece* 'lies' < \*leg'h-. The Faliscan development thus sets it off from all of Latinity, which is unified in this instance by its own shared innovation.

The conclusion to be drawn from this discussion is that features typically brought forth in favor of Faliscan as Dialectal Latin are inadmissible for determining the details of the genetic relationship of Faliscan to Latin; at best they show the results of independent changes or geographic diffusion of features. Moreover, if the other features discussed, concerning PIE \*g'h in various positions and concerning the verb for 'lie', are innovations not shared by Faliscan and Latin, then, by good dialectological criteria, Faliscan does not equate to some form of (Dialectal) Latin. Consequently, the model in (5) must be rejected in favor of the one in (6).

There are some morals to be drawn from all this discussion, ones that go beyond the microrelationships of Italic subgrouping and apply rather to methodology in general and to the question of macrorelationships. First, similarities alone are not enough to go on — a lot of careful sifting is needed to weed out the formal similarity of Latin *sum* and Oscan *súm*, for instance, and to focus in on the real shared innovations that underlie their formation. Also, similarities between later Faliscan and Latin (cf. (8)) are misleading; however tantalizing they seem, they give a false picture because they are chronologically off and do not come from the oldest available layer of Faliscan. In the end, with all the data to work from, just a few relatively reliable innovations emerged to lead to a conclusive determination about the relationship between Latin and Faliscan, but even those involve less than a handful of relevant forms. If this is

<sup>15</sup>As discussed in Joseph & Wallace 1991, the development of the preterite endings in Faliscan and Latin may be yet another example of the second type; it is not presented here as justifying it would involve more extensive discussion than the scope of the present paper permits.

what is available for a relatively well-documented group like Italic, how much more work will it take to get the best and most relevant facts for less thoroughly studied languages and language groups?

This concern is not mere idle stone-throwing. It is clearly the case that the shared innovation principle invoked herein cannot be used as a criterion for establishing a relationship between two groups, since shared innovation presupposes a relationship in the first place. However, a lot of what is discussed at the macro-level for language relationships, these days especially, really amounts to doing micro-level subgrouping. For example, Greenberg's claim (1987: 278) that there is an Almosan-Keresiouan group of languages is equivalent to saying that these languages form a subgroup within Amerind, yet among the evidence he cites is "the widespread occurrence of *s* as a second person marker". This "widespread occurrence" is nothing more than a shared similarity,<sup>16</sup> and no judgment is made of what is really the most crucial piece of establishing an Almosan-Keresiouan subgroup, namely whether this *s* is an innovation away from Proto-Amerind, and thus (possibly) significant for establishing such a subgroup, or instead is a shared retention, and thus inconclusive. It would seem that there is much to learn about macrorelationships from the examination of microrelationships, for the two pursuits are indeed related.

### References

- Bader, Françoise  
1976 "Le présent du verbe 'être' en indo-européen." *BSL* 71:17-111.
- Beeler, Madison  
1956 "Venetic and Italic." In: *Hommages à Max Niedermann*, ed. Georges Redard. Pp. 38-48. Bruxelles: Latomus.
- Beeler, Madison  
1963 "The Interrelationships Within Italic." In: *Indo-European and the Indo-Europeans*, eds. H. Birnbaum & J. Puhvel. Pp. 51-58. Berkeley & Los Angeles: University of California Press.
- Bomhard, Allan R. & John C. Kerns  
1994 *The Nostratic macrofamily: A study in distant linguistic relationship*. Berlin & New York: Mouton de Gruyter.

---

<sup>16</sup>See Rankin (1992: 340) for some discussion of problems with the Siouan data cited by Greenberg in this context.

Campanile, Enrico

- 1961 "Elementi dialetti nella fonetica e nella morfologia del Latino." *Studi e saggi linguistici* 1:1-21.

Campbell, Lyle

- 1988 Review article on Greenberg 1987. *Language* 64.3:591-615.

Eska, Joseph F.

- 1987 "The Language of the Latin Inscriptions of Pompeii and the Question of an Oscan Substratum." *Glotta* 65:146-161.

Giacomelli, Gabriella

- 1963 *La lingua Falisca*. Firenze: Leo S. Olschki.

Giacomelli, Roberto

- 1979 "Written and spoken language in Latin-Faliscan and Greek-Messapic." *Journal of Indo-European studies* 7:149-175.

Greenberg, Joseph H.

- 1987 *Language in the Americas*. Stanford: Stanford University Press.  
1989 "Classification of American Indian languages: A reply to Campbell." *Language* 65.1:107-114.

---

and Merritt Ruhlen

- 1992 "Linguistic origins of Native Americans." *Scientific American*, November 1992: 94-99.

Hock, Hans Henrich & Brian D. Joseph

- 1996 *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. (Trends in linguistics. Studies and monographs, 93.) Berlin and New York: Mouton de Gruyter.

Hoenigswald, Henry

- 1960 *Linguistic Change and Linguistic Reconstruction*. Chicago: University of Chicago Press.

Joseph, Brian D. & Rex E. Wallace

- 1987 "Latin *sum*, Oscan *súm*, *sim*, *esum*." *American Journal of Philology* 108:675-693.  
1991 "Is Faliscan a Local Latin Patois?" *Diachronica* 8.2:159-186.  
1992 "Socially determined variation in ancient Rome." *Language Variation and Change* 4:105-119.

- Leumann, Manu  
1977 *Lateinische Grammatik*. 1. Lateinische Laut- und Formenlehre. München: C. H. Beck'sche Verlagsbuchhandlung.
- Manaster Ramer, Alexis, Peter A. Michalove, Karen S. Baertsch, Karen L. Adams  
1997 "Exploring the Nostratic hypothesis." In: *Nostratic: Sifting the Evidence*, eds. Brian D. Joseph & Joseph C. Salmons. Amsterdam and Philadelphia: John Benjamins Publishing Co.
- Marinetti, A.  
1985 *Le iscrizioni sudpicene*. Firenze: Leo S. Olschki.
- Matisoff, James A.  
1990 "On megalocomparison." *Language* 66.1:106-120.
- Nichols, Johanna  
1992 *Linguistic diversity in time and space*. Chicago: University of Chicago Press.
- Pallottino, M.  
1968<sup>2</sup> *Testimonium Linguae Etruscae*. Firenze: La Nuova Italia.
- Palmer, Leonard  
1954 *The Latin language*. New York: Faber and Faber.
- Pisani, Vittore  
1962 "Il falisco nella formazione del più antico latino 'volgare'." *Ricerche linguistiche* 5:55-64.
- Pulgram, Ernst  
1978 *Italic, Latin, Italian: 600 B.C. to A.D. 1260*. Heidelberg: Carl Winter.
- Rankin, Robert L.  
1992 Review of Language in the Americas (Greenberg). *International Journal of American Linguistics* 58.3:324-351.
- Ross, Philip E.  
1991 "Hard Words." *Scientific American*, April 1991:138-147.
- Ruhlen, Merritt  
1994 *On the origin of language. Studies in linguistic taxonomy*. Stanford: Stanford University Press.

## Shevoroshkin, Vitaly

- 1989a "Methods in interphyletic comparison." *Ural-Altaische Jahrbücher* 61:1-26.
- 1989b (ed.) *Explorations in language macrofamilies*. Bochum: Universitätsverlag Dr. Norbert Brockmeyer.
- 1989c (ed.) *Reconstructing languages and cultures*. Bochum: Universitätsverlag Dr. Norbert Brockmeyer.
- 1990a "The mother tongue: How linguists have reconstructed the ancestor of all living languages." *The sciences*, May/June 1990: 20-27.
- 1990b (ed.) *Proto-languages and proto-cultures*. Bochum: Universitätsverlag Dr. Norbert Brockmeyer.
- 1991 (ed.) *Dene-Sino-Caucasian Languages*. Bochum: Universitätsverlag Dr. Norbert Brockmeyer.

## Sommer, Ferdinand

- 1977 *Handbuch der lateinischen Laut- und Formenlehre*. Heidelberg: Carl Winter.

## Vetter, Emil

- 1953 *Handbuch der italischen Dialekte*. Heidelberg: Carl Winter.

## Wallace, Rex E. &amp; Brian D. Joseph

- 1991 "On the Problematic *f/h* Variation in Faliscan." *Glotta* 69:84-93.
- 1993 "On the development of PIE \*g'h/gh in Faliscan: A response to Picard." *Diachronica* 10.1:144-150.

## Rigor or Vigor: Whither Distant Linguistic Comparison?

Mark Kaiser  
*University of California, Berkeley*

As my professor in three different seminars devoted to various topics in the field of historical linguistics and later in our scholarly collaborations, Vitaly Viktorovich Shevoroshkin repeatedly emphasized the importance of methodology in the reconstruction of protolanguages. In particular, he was skeptical of reconstructions not based on a rigorous correspondence of sound and meaning. Hence, his strong support for the Nostratic reconstructions of Dolgopolsky and Illič-Svityč was founded on their strict adherence to the comparative method.

Over the past several years a number of works have been published which fall under the general rubric of "distant linguistic comparison" or "deep reconstruction." These works have employed disparate methodologies, but the languages under investigation and at times even the titles of the works have been quite similar, which has led to some confusion, in particular inasmuch as a number of the works were written in Russian and translations are still unavailable. More recently, adherence to the comparative method's basic principle of a strict correspondence between sound and meaning has been under attack, and in one recent commentary was disparagingly dismissed as "Indo-baloney."

This recent research in distant linguistic relations, in particular the Nostratic Theory and Greenberg's (1987) *Language in the Americas*, has enlivened the field of historical linguistics, if in no other way by forcing us to reexamine our basic methodologies. Amidst a fair amount of confusion, two positions have formed: those who reject out of hand any claim of genetic affinity which fails to meet the degree of proof established by Indo-European, and those who randomly pick out stems which are only **similar** in sound and meaning and, without reconstruction of the proto-language, claim genetic cognates.

Unfortunately, the discipline has been polarized by this new research. The "rigorists" reject all scholarship in distant linguistic relations without regard to the methodology employed, and "vigorists" seem too willing to accept all work in distant linguistic comparison as equally valid.

This paper entails a survey of the methodologies employed in distant linguistic comparison over the past two decades. I can only hope that Vitaly Viktorovich will find herein traces of his exhortation that "methodology matters."

### The Moscow School

Although theirs was the first work to appear on the scene (in various Russian journals and series in the mid-1960s and in the first two volumes (1971, 1976) of Illič-Svityč's *Nostratic Dictionary*, Dolgopolsky's and Illič-Svityč's

work remains mostly unknown in the U.S. today. Their Nostratic theory, developed independently, reconstructs the protolanguage (Nostratic) which gave rise to Afro-Asiatic (= A-A), I-E, Kartvelian, Altaic, Uralic, and Dravidian (see Addendum I - a few translated entries from Illič-Svityč 1971 (henceforth, Illič-Svityč 1971, 1976, and 1984 = *Nostratic Dictionary*). The Nostratic reconstructions are based on strict adherence to the comparative method, i.e., on a regular system of phonetic correspondences. The *Nostratic Dictionary* is a cautious, conservative work, which is reflected in the Russian title: *Opyt sravnenija nostratičeskix jazykov* [*An Experiment in Comparison of the Nostratic Languages*]. Throughout, Illič-Svityč hedges his own claims, places some etymologies under question, and suggests alternative interpretations.

Illič-Svityč and Dolgopolsky's early work in the 1960s provided the theoretical and methodological foundation for their later Nostratic reconstructions. Illič-Svityč 1964a demonstrates the importance of distinguishing lexical borrowings from genetic cognates. Illič-Svityč notes that among Möller's "fantastic and implausible" comparisons between Semitic and Indo-European, there is a small group of cognates which, as cultural items, are unlikely to form the basis of a Semitic-Indo-European proto-language, as Möller had assumed. He shows that these cognates, due to the nature of their root structure, can best be explained as borrowings from Semitic into I-E. Moreover, the presence of good etymologies in A-A and their absence in I-E also supports this interpretation. Illič-Svityč's interest in borrowing may also be seen in Illič-Svityč 1965, where examples of borrowing to and from Caucasian languages (North and South) are deduced.

Dolgopol'skij 1964a was originally intended as a guide for determining whether **further** etymological study of a group of languages is warranted. He begins with two premises: 1) More than one language (language family) must be used in comparison (when only two languages are compared, the probability that the results will be contaminated by chance phonetic similarities is high); and 2) Not all lexemes are suitable for comparison; in fact we are interested in those morphemes which for the given group of languages show a high degree of stability. He then demonstrates that certain morphemes are more stable, that there exists a hierarchy of lexical stability, i.e., certain lexemes are less likely to be replaced by other stems with the same meaning. Next, he divides phonemes into 10 groups according to their likelihood of diachronic correspondence, i.e., *t, d, d<sup>h</sup>, t<sup>h</sup>, ʈ* fall into one group (T), *p, b, f, p* into another (P), etc. He examines the stems of his most stable lexemes in I-E, A-A, Uralic, Altaic, Chukchee-Kamchatkan, Kartvelian, and Sumerian, rewrites them in terms of his 10 groups of phonemes, and concludes that statistical analysis precludes the possibility that these are chance correspondences (except in the case of Sumerian). In certain respects this study resembles the methodology of "multilateral comparison" (see below) in that Dolgopolsky is not concerned with specific sound correspondences, but is trying to present broad similarities



between language groups. However, Dolgopolsky differs from the mass comparatists in that he considered this only a preliminary step (or a test of plausibility of further study) and followed up with a reconstruction of precise phonetic correspondences (see bibliography).

Thus, both Illič-Svityč and Dolgopolsky took the possibility of borrowing into account and in some instances used borrowing to explain resemblances between some words. However, borrowing was rejected as an explanation for a larger corpus of data. As we examine their articles chronologically, we can follow the development of their methodology: Dolgopolsky's general juxtaposition of phonetically similar items described above was followed by the first attempts to reconstruct the proto-language (Illič-Svityč 1964b), where it was pointed out that I-E labial velars corresponded to Uralic *\*kU* (where *U = o, u or ü*), I-E palatal velars corresponded to Uralic *\*kE* (*E = e, i, ä*), and I-E plain velars corresponded to Uralic *\*ka*. Dolgopol'skij 1964c presented 166 sets of lexical correspondences between I-E, Uralic, Altaic, A-A, Kartvelian, and Chukchee-Kamchatkan. The material was arranged according to regular phonetic correspondences, but no attempt at reconstruction of the proto language was made. Many of these sets later became part of the *Nostratic Dictionary*, some were rejected, and others reinterpreted. For example, for set #152 (I-E *\*kerd-* 'heart' ~ Kart. *\*mkerd-* 'breast' ~ Sem. *\*krb* 'viscera, thorax') the comparison of the I-E ~ Kartvelian data was retained, but the Semitic data were incorporated into another root, resulting in Nst. *\*\*kErdV* 'heart' [Illič-Svityč 1971:324-5] and *\*\*Karbli* 'abdomen, viscera' [Illič-Svityč 1971:338-40].

The first reconstruction of the Nostratic proto-language was given in Illič-Svityč 1967, where the data are arranged according to reconstructed semantics (Russian alphabetic order). Tables of phonetic correspondences for all phonemes (not just stops) and 607 sets (the number of reconstructed Nostratic roots is slightly less) are provided. Within each set each family's proto-form is reconstructed and bibliographic information is furnished. Slightly more than half of these comparisons involve data from only two languages, and on average 2.75 language families are involved in each comparison. At this point one may speak of the "canonization" of Nostratic in that the languages under comparison and phonetic correspondences become fairly well established, although a number of changes are introduced in the first volume of the *Nostratic Dictionary* (Illič-Svityč 1971) and research in the 1980's has further modified the Nostratic phonological system.

The *Dictionary* is a more conservative and complete work. Although there are more roots in Illič-Svityč 1967 (the three volumes of the *Dictionary* published to date contain 378 entries), the *Nostratic Dictionary* provides the reconstructed form for each language family, the evidence from individual languages used in the reconstruction, and extensive commentary on incongruities. It is clear that the *Dictionary* maintains a more cautious approach

to reconstruction. Of the 36 words in *b-* found in Illič-Svityč 1967, 26 were included in the *Dictionary* (and six new roots were added). Nine of the ten excluded items were binary comparisons, which suggests that their genetic affinity has not necessarily been rejected, but simply not yet proven. In the *Dictionary* more than 63% of the entries entail comparison of more than two language families and on average 2.98 language families are involved in each comparison. And finally, the *Dictionary* is not limited to the lexicon: Illič-Svityč reconstructed seven personal or demonstrative pronouns, eighteen affixes, including diminutive, plural, comparative degree, etc., and ten particles, including negation, incitement to action, locative, etc.

The *Dictionary*, and this can be said of Nostratics in general, is a work very much "in progress." Data from Illič-Svityč 1967 underwent a number of modifications before inclusion in the *Dictionary*: some phonemes were dropped, others added, reflexes were changed. In many instances data from a language family not represented in Illič-Svityč 1967 were added, in some cases data from a family were rejected. In the *Dictionary* Illič-Svityč rejected the notion of Nostratic roots with a vowel in anlaut, preferring to reconstruct initial laryngeals which were later lost in all the East Nostratic languages (Altaic, Uralic, Dravidian), and whose fate in West Nostratic (I-E, A-A, Kart.) depended on language family and type of laryngeal. Perhaps the most significant difference is the addition to the *Dictionary* of over 35 grammatical suffixes and pronouns.

Dolgopolsky continued to publish articles on Nostratic problems, but with a much narrower focus. For example, Dolgopol'skij 1969 and 1972 are concerned with the reflexes of Nostratic consonant clusters and Dolgopol'skij 1974 deals with the phoneme /z/ in Nostratic. In addition, Dolgopolsky assisted Dybo in editing the manuscript of Illič-Svityč 1971 and published a reconstruction of proto-Cushitic. Dolgopolsky 1984 deepens our comprehension of Nostratic grammar by examining the system of Nostratic personal pronouns. In this work, in many respects an extension and further elaboration of Illič-Svityč 1971 and 1971b, he not only describes the evolution of personal pronouns from Nostratic to the various daughter languages, he also is able to deduce a number of Nostratic syntactic rules. Three basic types of words (full words, pronouns, grammatical words) are identified for Nostratic, as well as a Subject-Object-Verb word order.

After Dolgopolsky's emigration to Israel, the Nostratic tradition fell on the shoulders of V. Dybo, who had edited the publication of all three volumes of Illič-Svityč's *Dictionary*. Dybo conducted the Illič-Svityč Seminar on distant linguistic relations at the Institute of Slavic and Baltic Studies in Moscow. His students participated in the preparation of volume three of the *Dictionary* (1984), and then moved on to work on the reconstruction of other language families, including Northeast and Northwest Caucasian, Sino-Tibetan, and Dene-Caucasian (see Nikolaev and Starostin 1964). The methodology they employed has been in

the tradition of Illič-Svityč and Dolgopolsky, the principles of which are worth reiterating:

- 1) Comparison of multiple languages or language families - binary comparisons are to be avoided.
- 2) Reconstruction of the protolanguage by means of a system of strict phonological and semantic correspondences.
- 3) The possibility of borrowing must be taken into account.
- 4) Treatment of the data involves ongoing modification and refinement.

At an early stage, Illič-Svityč's and Dolgopolsky's articles were similar to the mass comparison technique of J. Greenberg.

### **Nostratics in America**

Vitaly Viktorovich has spent an extraordinary amount of time and energy propagating Illič-Svityč and Dolgopolsky's accomplishments in historical linguistics. The rewards have been few, the frustrations many. He has struggled against the anti-historical bias in American linguistics and against fellow Indo-Europeanists who were unwilling to consider any proposal of a genetic relationship involving Indo-European and other language families. At the same time he made contributions to Nostratic theory, including modifications of the reconstruction of Nostratic laryngeals, pharyngeals, and post-velar stops and their reflexes in I-E (Kaiser-Shevoroshkin 1985; Shevoroshkin 1988a), the use of Nostratic Theory and borrowings to refute the theory of glottalized stops for I-E and to propose an I-E system of tense, lax, and voiced (Shevoroshkin 1986), and the reconstruction of Nostratic laterals (Kaiser-Shevoroshkin 1988). His efforts have been complicated by the work of A. Bomhard, who introduced his own version of Nostratic, differing in fundamental respects from the work of the Moscow school.

Bomhard 1984 is a binary comparison of I-E and A-A (primarily Semitic) data. We have already noted Illič-Svityč and Dolgopolsky's reluctance to work with binary comparisons. Binary comparisons are inherently limited: the more languages involved in a reconstruction, the more controls on speculation. If we were to reconstruct I-E on the basis of only two of the language groups (e.g., Germanic and Greek), our reconstruction would differ significantly from what we now posit for I-E. In a binary comparison, it is more difficult to filter out chance similarities or to distinguish borrowings from genetic cognates.

Bomhard 1984 makes many important contributions to the Nostratic corpus. However, in addition to being primarily a binary comparison, there are numerous other methodological errors in the work: roots are truncated or in other ways modified to better match data from other languages, and the semantics of reconstructions are embellished (for details, see Kaiser-Shevoroshkin 1987, Palmaitis 1986). Other scholars duly noted the lack of a rigorous methodology and dismissed not only Bomhard's questionable reconstructions, but, unfortunately, his good reconstructions and the entire concept of Nostratics, as

well.

In Bomhard 1992 we find a general discussion of Nostratics, where Bomhard includes Indo-European, Afro-Asiatic, Kartvelian, Uralic, Dravidian, Altaic, Chukchi-Kamchatkan, Gilyak, Eskimo, and possibly Sumerian. He also provides examples of personal pronouns (although this topic is more thoroughly treated in Dolgopolsky 1984, which is not cited) and tables of phonetic correspondences (differing in many respects from Illič-Svityč 1971). Supporting data for his phonetic correspondences are presented in Bomhard and Kerns (1994).

One major difference between Bomhard, on the one hand, and Illič-Svityč and Dolgopolsky, on the other, is the treatment of what is traditionally reconstructed as I-E voiced stops. Bomhard follows the Hopper-Gamkrelidze-Ivanov analysis of I-E consonantism and reconstructs a glottalized series for I-E. This glottalized series is compared with glottalized series in Kartvelian and A-A and a glottalized series is thereby reconstructed for Nostratic, whereas Illič-Svityč compared I-E voiced stops to voiceless stops in Kartvelian and A-A. Thus, in somewhat simplified form:

<u>Per Bomhard:</u>	A-A	Krt	I-E	<	Nost.
	T	T	T	<	T
	D	D	D <sup>h</sup>	<	D
	ṭ	ṭ	ṭ	<	ṭ

<u>Per Illič-Svityč:</u>	A-A	Krt	I-E	<	Nost.
	T	T	D	<	T
	D	D	D <sup>h</sup>	<	D
	ṭ	ṭ	T	<	ṭ

(Note: Bomhard's I-E *ṭ* – Illič-Svityč's I-E *D*, i.e., for both scholars these generate the same reflexes in the I-E daughter languages < T – a voiceless stop, D – a voiced stop, D<sup>h</sup> – a voiced aspirated stop, ṭ – a glottalized stop).

The issue is not whether we should reconstruct a glottalized series for I-E (although there are strong arguments **against** such a reconstruction, see Kaiser-Shevoroshkin 1986). The problem for Nostratics is which of the two correspondences is correct: for example, do we follow Illič-Svityč and compare A-A roots in *ḳ* and Kartvelian roots in *ḳ* with I-E roots in *k* or do we follow Bomhard and compare them with I-E roots in *g* (old notation) = *ḳ* (the "new" notation per Bomhard, Hopper, et al.)? We must wait for Bomhard-Kerns (1994) to examine the evidence supporting his contention. In the meantime we can only conclude that that evidence is lacking in Bomhard's writings to date.

## The Methodology of Multilateral Comparison

The methodology of multilateral comparison is described and defended in Greenberg 1987:1-37 and Ruhlen 1987:120-124. Data from many languages are gathered together into semantic groups and then further divided according to broad phonological similarities, but no system of regular sound correspondences is established. The sole purpose of the methodology is to establish genetic relationships, not to reconstruct protoforms of the languages. The question that remains, however, is whether multilateral comparisons are sufficient to establish genetic affinity. Greenberg's book begs for elaboration. For example, for the sememe 'woman' the forms *čalo-na*, *kilaua*, *kelaa*, *kila*, *kili-p* from various languages of Macro-Panoan are compared to the forms *kvantua*, *kneu*, *\*kuja*-, *kunja*, *igūn* from languages of Equatorial Amerind. Superficially, this comparison appears reasonable, but on further scrutiny one is left with too many unanswered questions. It would seem that in the first case we are dealing with the reflexes of *\*kEl-* and in the second with *\*ku(j)n-*, but we cannot be certain, because there are no reconstructions provided. In any event there are no other examples in Greenberg's text of a correspondence between Macro-Panoan *-l-* and Equatorial *-n-* or *-jn-* or *-nj-*, and without that regularity it is too easy to come to the conclusion that this is a chance sound resemblance. This is compounded by liberties taken in the semantic comparisons: 'feather' ~ 'leaf', 'strong' ~ 'bone', 'small' ~ 'daughter', 'light' ~ 'burn' ~ 'sun', 'burn' ~ 'star', 'shoulder' ~ 'arm' ~ 'back'. None of these comparisons is objectionable by itself; however, coupled with the lack of sound correspondences, the reader is left unconvinced.

Greenberg's approach is a valuable first stage in determining which languages need be compared, but it must be rejected as a method of proof of genetic affinity. First, it is impossible to distinguish archaic borrowings from true genetic cognates, and second, we can never with full assurance dismiss chance similarities as an explanation. If we knew nothing of the history of English and were to mistakenly assume that English also constitutes a family within Amerind, there are numerous examples where an English word approximates forms in other Amerind languages: in the given set, *queen* would match up well with the Equatorial forms. If instead of English we were to add all Germanic languages, the number of forms able to match something in Amerind would grow dramatically. In the absence of the control of regular sound correspondences, each additional language group increases the number of cognates geometrically.

Greenberg's work is an important step in the process of reconstruction: it gives the linguist an idea of which languages fit where before application of the comparative method, but it is not an end in itself. And once regular phonetic correspondences have been established, then chance must be rejected as an explanation for similarities. Regular phonetic correspondences between languages, reconstructed language groups, or reconstructed language families can be explained only as a result of common genetic origin or borrowing, and languages simply don't borrow the formation of the past tense as a result of

contact, nor is the entire lexicon subject to borrowing. Greenberg's methodology has also been used in the so-called global etymologies, where the attempt is made to prove a genetic relationship across macrofamilies. Thus, in the example above, it has been claimed that the Equatorial forms are indeed related to English *queen* < I-E \*g<sup>w</sup>en- < Nostratic \*küni (see appendix). These types of comparisons are exceptionally fascinating and exceedingly premature. Before these comparisons can be seriously entertained, the macrofamilies involved need to be fully reconstructed, otherwise the comparisons will be suspect as nothing more than chance similarities.

### A New Approach To Proof of Genetic Relationship

Starostin 1991 describes a new methodology to establish a genetic affinity between languages. Starostin's goal is to demonstrate the place of Japanese within the Altaic language family, but first he must establish the existence of an Altaic family. He begins by constituting the regular sound correspondences within the Altaic branches (Turkic, Mongolian, Tungus, and Korean). He then uses a modified version of Swadesh's 100-word list ('all', 'bark', 'bite', 'feather', 'flow', 'lie', 'nail', 'seed', 'warm', 'we' are rejected and replaced with 'far', 'heavy', 'near', 'salt', 'short', 'snake', 'thin', 'wind', 'worm', 'year') and, where available, provides Altaic etymologies for these sememes. The 100-word list is divided into two, one 35-word list of most stable lexemes ('blood', 'bone', 'die', 'dog', 'ear', 'egg', 'eye', 'fire', 'fish', 'full', 'give', 'hand', 'horn', 'I', 'know', 'louse', 'moon', 'name', 'new', 'nose', 'one', 'salt', 'stone', 'sun', 'tail', 'this', 'thou', 'tongue', 'tooth', 'two', 'water', 'what', 'who', 'wind', 'year') and the remaining 65 words. If the percentage of cognates from the 35-word list is greater than in the 65-word list, it means that the cognates are genetic in origin and not chance resemblances. For example, of the 100 words of the modified Swadesh list, there are 16 correspondences between proto-Turkic and proto-Tungus, eight of which are in the 35-word list (22.9%) and eight in the 65-word list (12.3%). Thus, this method permits us to eliminate the possibility that our reconstructions are only chance resemblances between words, because words that are more stable in the lexicon show a significantly higher percentage of correspondence.

It should be clear from this discussion that the dichotomy of "rigor" and "vigor" is a choice we need not make: we can have our cake and eat it, too. The work of Illič-Svityč, Dolgopolsky, and more recently Starostin has provided new vigor to historical linguistics while at the same time maintaining high standards. Their methodology is within the scope of the comparative method, i.e., it is based on rigorous observation of regular phonetic correspondences. Nor can the multilateral comparisons or global etymologies be rejected out of hand: these are the **first steps** in the process of establishing a genetic relationship between languages and the reconstruction of their mother tongue.

## References

- Bomhard, Allan R.
- 1984 *Toward Proto-Nostratic: A New Approach*. Current Issues in Linguistic Theory, 27. Amsterdam: John Benjamins.
  - 1991 Lexical parallels between Proto-Indo-European and other languages. *Studia etymologica Indoeuropaea: Memoriae A. J. Van Windekens (1915-1989) dicata*. Ed. L. Isebaert. Leuven: Peeters.
  - 1992 The Nostratic macrofamily (with special reference to Indo-European). *Word* 43(1), 61-83.
- Bomhard, Allan R. and John C. Kerns
- 1994 *The Nostratic Macrofamily: A Study in Distant Linguistic Relationship*. Berlin and New York: Mouton de Gruyter.
- Dolgopol'skij, A. B.
- 1964a Gipoteza drevnejšego rodstva jazykovyx semej Severnoj Evrazii s verojatnostnoj točki zrenija. *VJa* 2:53-63. For English translation see Dolgopolsky 1986.
  - 1964b Metody rekonstrukcii obščeeindoevropskogo jazyka i vneindoevropskie sopostavlenija. *Problemy sravnitel'noj grammatiki indoevropskix jazykov. Tezisy dokladov*. Moscow: Nauka.
  - 1964c Gipoteza drevnejšego rodstva jazykov severnoj Evrazii (problemy fonetičeskix sootvetstvij). *VII Meždunarodnyj kongres antropologičeskix i ètnografičeskix nauk*. Moscow: Congress publications.
  - 1969 Nostratičeskie osnovy s sočetaniem šumnyx soglasnyx. *Ètimologija* 1967. Moscow: Nauka. 296-313.
  - 1971 Nostratičeskie ètimologii i proisxoždenie glagol'nyx formantov. *Ètimologija* 1968. Moscow: Nauka. 237-242.
  - 1972 Nostratičeskie korni s sočetaniem lateral'nogo i zvonkogo laringala. *Ètimologija* 1970. Moscow: Nauka. 356-369.
  - 1973 *Sravnitel'no-istoričeskaja fonetika kušitskix jazykov*. Moscow: Nauka.
  - 1974 O nostratičeskoj sisteme affrikat i sibiljantov: korni s fonemoj \*3. *Ètimologija* 1972. Moscow: Nauka. 163-175.
- Dolgopolsky, Aaron B. (=Dolgopol'skij, A.B.)
- 1984 On personal pronouns in the Nostratic languages. *Linguistica and Philologica. Gedenkschrift für Björn Collinder*. Vienna.

- 1986 A probabilistic hypothesis concerning the oldest relationships among the language families in northern Eurasia. *Typology, Relationship and Time*. V. V. Shevoroshkin and T. L. Markey, eds. Ann Arbor: Karoma. 27-50.

Gamkrelidze, T.V., and Vjačeslav V. Ivanov

- 1984 *Indoevropskij jazyk i indoevropcejcy: Rekonstrukcija i istoriko-tipologičeskij analiz prajazyka i protokul'tury*. Tbilisi: University of Tbilisi.

Greenberg, Joseph H.

- 1957 Genetic relationship among languages. *Essays in Linguistics*. Chicago: U of Chicago Press.
- 1987 *Language in the Americas*. Stanford: Stanford University Press.

Illič-Svityč, V. M.

- 1964a Drevnejšie indoevropsko-semitskie jazykovye kontankty. *Problemy indoevropskogo jazykoznanija*. Moscow: Nauka. 3-11.
- 1964b Genezis indoevropskix rjadov guttural'nyx v svete dannyx vnešnego sravnenija. *Problemy sravnitel'noj grammatiki indoevropskix jazykov*. Moscow: Nauka. 22-26.
- 1965 Caucasic. *Ėtimologija 1964*. Moscow.
- 1967 Materialy k sravnitel'nomu slovarju nostratičeskix jazykov. *Ėtimologija 1965*. Moscow: Nauka. 321-373.
- 1968 Sootvetstvija smyčnyx v nostratičeskix jazykax. *Ėtimologija 1966*. Moscow: Nauka. 304-405.
- 1971 *Opyt sravnenija nostratičeskix jazykov (semitoxamitskij, kartvel'skij, indoevropskij, ural'skij, dravidijskij, altajskij). Vvedenie. Sravnitel'nyj slovar' (b-ŕ)*. Moscow: Nauka.
- 1971b Ličnye mestoimenija mi 'ja' i mä 'my' v nostratičeskom. *Issledovanija po slavjanskomu jazykoznaniju. Sbornik v čest' šestidesjatiletija professora S. B. Bernštejna*. E. B. Češko, et al., eds. Moscow: Nauka.
- 1976 *Opyt sravnenija nostratičeskix jazykov (semitoxamitskij, kartvel'skij, indoevropskij, ural'skij, dravidijskij, altajskij). Sravnitel'nyj slovar' (l-ž)*. Moscow: Nauka.
- 1984 *Opyt sravnenija nostratičeskix jazykov (semitoxamitskij, kartvel'skij, indoevropskij, ural'skij, dravidijskij, altajskij). Sravnitel'nyj slovar' (p-q)*. Moscow: Nauka.



Kaiser, M., and V. Shevoroshkin

- 1985 On Indo-European laryngeals and vowels. I. Laryngeals before vowels. *JIES* 13(3-4):377-413.
- 1986 Inheritance versus borrowing in IE, Kartvelian, and Semitic. *JIES* 14(3-4):365-378.
- 1987 On recent comparisons between language families: The case of Indo-European and Afro-Asiatic. *General Linguistics*, 27(1):34-46.
- 1988 Nostratic. *Annual Review of Anthropology* 17:309:329.

Möller, Hermann

- 1907 *Semitisch und Indogermanisch*. Kopenhagen
- 1911 *Vergleichendes indogermanisch-semitisches Wörterbuch*. Reprinted 1970. Göttingen.

Nikolaev, S. L., and S. A. Starostin

- 1984 Severnokavkazskie jazyki i ix mesto sredi drugix jazykovyx semej Perednej Azii. *Lingvističeskaja rekonstrukcija i drevnejšaja istorija vostoka: Tezisy i doklady konferencii*. Moscow: Nauka. 3:26-34.

Palmaitas, Mykolas L.

- 1986 New Contributions to 'Proto-Nostratic'. (Review of Bomhard 1984). *IF* 91:305-317.

Ruhlen, Merritt

- 1987 *A Guide to the World's Languages, Vol. 1: Classification*. Stanford: Stanford University Press.

Shevoroshkin, V. V.

- 1988a On laryngeals. *Die Laryngalthetheorie und die Rekonstruktion des indgermanischen Laut- und Formensystems*. Heidelberg: Carl Winter.
- 1988b Indo-European consonants in Anatolian. *A Linguistic Happening in Memory of Ben Schwartz. Studies in Anatolian, Italic, and Other Indo-European Languages*. Yoël L. Arbeitman, ed. 283-203.
- 1989 Methods in Interphyletic comparison *Ural-Altaische Jahrbücher* 61:1-23.

Shevoroshkin, V. V., and Thomas L. Markey, eds.

- 1986 *Typology, Relationship, and Time: A Collection of Papers on Language Change and Relationship by Soviet Linguists*. Ann

Arbor: Karoma.

Starostin, S. A.

- 1982 Praenisejskaja rekonstrukcija i vnešnie svjazi enisejskix jazykov. Ketskij sbornik: Antropologija, ètnografija, mifologija, lingvistika. E. A. Alekseenko, et al., eds. Leningrad: Nauka. 144-238.
- 1984 Gipoteza o genetičeskix svjazjax sinotibetskix jazykov s enisejskimi i severnokavkazskimi jazykami. Lingvističeskaja rekonstrukcija i drevnejšaja istorija vostoka: Tezisy i doklady konferencii. Moscow: Nauka. 4:19-38.
- 1985 Kul'turnaja leksika v obščeseverokavkazskom slovarnom fonde. Drevnjaja Anatolija. B.B. Piotrovskij, et al., eds. Moscow: Nauka. 74-94.
- 1987 Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy. Moscow: Nauka.
- 1991 *Altaiskaja problema i proisxoždenie japonskogo jazyka*. Moscow: Nauka.

## APPENDIX

SAMPLE RECONSTRUCTIONS TRANSLATED FROM ILLIČ-SVITYČ'S  
NOSTRATIC DICTIONARY

(Note: S-H = Semito-Hamitic = Afroasiatic)

#65. *-di*, suffix of past tense forms: Kartvelian *\*-di*, suffix of the imperfect ~ Dravidian *-tt/-t-*, suffix of the preterit ~ Alt. *-di*, suffix of the preterit.

Krt.: suffix of the imperfect: Georgian *-di* (1-2 ps.; in 3rd ps. sg. *-da* < *\*di-a*) || Megrel *-di* (1-2 ps., in 3rd sg. *-du* with loss of *\*i* before the marker of the 3rd sg., *-u* < *\*-a*); in Chan *-ti-*, where *t* < *d* by dissimilation (cf. Zhgenti Chan 140 for similar cases) || Svan *-d*; the proposal by Deeters 135 that the Svan formant was borrowed from Georgian is unwarranted. || Cf. Klimov 65 (where *\*-d* is reconstructed).

Drv.: suffix of preterit: Tamil. *-tt/-t-* (depending on the stem ending, which probably reflects the state of affairs in Proto-Dravidian. Cf. *cej-t-e:n* 'I made', but *paṭi-tt-e:n* 'I learned'); Tulu *-t* (< *\*-tt*); Kannada *-d* (< *\*-t*); Tulu *-t/-d-* || Central Drav.: Kolami, Parji *-t/-d-*, Naiki *-t-*, Gondi, Konda *-t-*, Kui *-t-* || Kurux, Malto *-t/-d-* || See Emeneau TPhS 1957, 36-43; Bloch 53-7; Andronov 71-2.

Alt.: Turkic *\*-ði/-ḡy* and *\*-ti/-ty* (the second variant after *r, l, n*), suffix of preterit, see Räsänen Morph. 229-230. The proposal (Pope Islamica 1, 424) that this suffix contained the marker of the 3rd sg. *\*i* is implausible. || Mongolian *\*-ṣi / -či* (< *\*-di*; *ṣi* > *-či* after *b, d, g, s, r*) and the secondary variant *\*-ṣu / -ču*, suffix of converbium imperfecti: Middle Mongolian (MA) *-ṣi / -či*, WrMongol. *-ṣu / -ču*, Dagur, Khalkha *-ṣi / -či*, Mongor *-ṣi*. This suffix is included in the formant of the preterit *\*-ṣuGuil-ṣügüi* || Tungus: *\*-dal-dä*, suffix of aorist ("present tense") of class-III verbs, cf. 1st sg. aor. from *\*ga-* 'to take, buy': Ulcha *ga-da-mbi* (2nd present), Evenki, Even *ga-da-m* and others (Sunik Glag. 77-8 incorrectly considers *\*-da* a phonetic variant of the suffix *\*-ra*) || ? MidKorean *-id-* (modern *-at/-t-*), a formant of the preterit (see Boo Kyom 94) || Cf. Pope CAJ 2,204; Ramstedt MSFOu 19,106-7, Benzing UAJb 24,131-2; Räsänen Morph. 229-30; Pope Mong. 277,265-6; Benzing 1074-5. The variant *\*-di*, represented in Turkic and Mongolian, is most likely original; in Tungus, *\*-dal-dä* can be assumed to be leveling with suffixes of the I and II verbal classes *\*-ral-rä* and *\*-*

*sa/-sä.*

? I-E: It is worth mentioning Germanic preterit suffixes of weak verbs *\*-da* (cf. 1st. sg. pret. Goth. *lagi-da*, OIc. *lagða*, OHG. *legi-ta* 'laid'), which have not received a convincing etymology (tying them to I-E. *\*d<sup>h</sup>eh-* 'to put' or to I-E-*t-* pose significant difficulties. Cf. Prokosch 204-10).

- ◆ Cf. Cald. 510 (Altaic ~ Dravidian). The original *\*i* is preserved in Altaic and Kartvelian; in Dravidian *\*-tt-/-t-* with the loss of cerebralization as a result of the very archaic devoicing of *d*, which was in auslaut after the loss of the vowel. The archaic meaning of the formant was probably purely temporal: even in Kartvelian, where the verbal system is dominated by aspectual contrast of the present to the aorist, the imperfect in *\*-di* preserves a purely temporal meaning.

#178. *küni* 'wife, woman': S-H *k(w)n / knw* 'one of the wives (in polygamy), woman' ~ I-E. *g<sup>u</sup>en-* 'wife, woman' ~ Alt. *küni* 'one of the wives (in polygamy)'.

S-H: Semitic: Akkad. (Late Babylonian) *kini:tu*, (New Akkad.) *kini:tuu* | *qini:tu* f., probably 'one of the wives (in polygamy), female friend in the harem' (Soden AW 480; Muss 410 has 'servant'). In Aramaic we see a semantic shift from a feminine noun to male individuals: Bibl.Aramaic *kənāwā te:* pl. 'friends', OAram. *knt*, Syriac *kəno: to:* 'friends' 2 Berber *\*t-knw* f.: Tuareg *te:kne* (pl. *te:knewi:n*), Sus *takna* (pl. *takniwin*), Kabil *takna* 'one of the wives in polygamy (in relation to other wives)'. Apparently, the corresponding masculine form with the meaning 'twin' is a secondary development: Tuareg *e:kne* (pl. *e:knewen*), Sus *ikennu* (pl. *iknuan*) || Cushitic *\*H-kwn* with the prefix *H(V)-* (the cluster *\*Hk-* is explained by the development of *g*- and *q*- in anlaut in numerous Cushitic languages); Central Cushitic 'woman': Bilin (Reinisch) *q̄gi:na:* (pl. *uḱ<sup>w</sup>i:n*) [Ed. note: according to Palmer 'əx<sup>w</sup>ina (pl. 'əḱ<sup>w</sup>in)], Khamir *iu:na:* (pl. *uḱūn*), Khamta *eq<sup>w</sup>en*, Dembya *kiu:na:* (pl. *k<sup>u</sup>i:n*), Kemant *jiwi:na:*, Kuara *iewi:na*, Aviiia *xuona:* Galla *gena* 'lady, legal wife of king'; West.Cushitic: Chara *gāne:ts* 'woman', Kaffa *genne* 'lady' (queen's title), Mocha *gänne* 'lady, woman', Shinasha (Beke) *genna*, (d'Abbadie) *žänna:* 'lady' || ? Chadic: Chibak *ḡkwā*, Margi *ḡkwā* 'girl' (*ḡk-* < *\*m-k-* with prefix *m-*); Kotoko *\*ḡgen-*, *\*ḡgenVm* (and later *ḡgerVm* with dissimilation; possibly, *ḡg-* < *\*m-k-*) 'woman': Ngala (Migeod) *ginum* (von Duisburg *genim*), Makari, Affade *gerim*, Shoe (Koelle) *ḡgeram*, Kusri *gerum*, Gulfei

*gəram*, Logone *gənam* (stem *gən-* with possessive suffixes: *gən-tu-'u* 'my wife' and similar cases, see Lukas Log. 30), Kuri (*i*)*ŋgerim*, Buduma *ŋgərùm* || Cf. Rössler ZAss 50, 133; Ges. 910; Reinisch Chamir 106, 25; Cerulli St. 3, 168; Cerulli St. 4, 445; Conti Rossini RStO 6, 408; Sölken Anthropos 53, 893. In Semito-Hamitic there are two variants: *\*kwn* (Cushitic, possibly Chadic) and (with metathesis) *\*knw* (Semitic, Berber). The shift of S-H. *\*k(w)n* to male individuals in Semitic and Berber derives from the period of existence of a punalual family structure, when the corresponding word signified 'a woman from the conjugal class of wives'; the masculinization of this form gave the meaning 'man from the conjugal class of husbands' (whence later 'twin, friend').

I-E: 'wife, woman': OI. *gna*: ('goddess, woman of divine origin'; see Mayr. 1, 351 for traces of archaic disyllabicity); Av. *gəna*: || Arm. *kin* (pl. *kanai-k'*) || Grk. *γυνή* (gen. *γυναικός*); Myc. *ku-na-ja*, probably = *γυνή* 'female (adj.)', Cf. Morpurgo 168), Boeotic *βανᾶ* || Alb. (Gheg) *grue*, (Tosk) *grua*, (< *\*g<sup>u</sup>n-o:n*) || Messapian *benna* || OIr. *ben* (gen. sg. *mná*) || Goth. *qino*, OHG *quena* 2 OPruss. *genno*; OCS *žena* || Toch. A *śām* (pl. *śnu*), B *śana* (obl. *śno*) 2 See Pok. 473-4.

Alt.: Turk. *\*küni* (*k* is indicated by Oguz *g-*) 'one of the wives in polygamy (in relation to the other wives)': OTurkic (Yeniseyan) *küni*, OUighur *küni* (cf. *kün-tä-ki* 'located in the female half' - a derived adjective in *-ki* from the locative in *-tä*); Kirghiz *künü*, Uzbek *kundoš*, Bashkir *kondäš* (compounded with *\*-daš* 'friend' - see Ramstedt SKE 257), OKypchak *küni*; Turkmen *güni*, Azerbaijani *günü*, Turkish (Erzrum, see SD 688) *günü* ('friend'). See Pokrovskaja IRLTJa 66. The suppositions that the Turkic word was borrowed from SinoKorean *hu-nje* 'harem wife' (Ramstedt SKF, 65) or that it is related to Turkic *\*k'ün* 'jealousy' are in error (Egor 122; cf. #229 above).

- ♦ Cf. Trombetti 66 (I-E ~ Turkic). The *ü* vocalism in the first syllable, preserved in Altaic, is indirectly reflected in S-H. (*\*w*) and I-E. (labiovelar *\*g<sup>w</sup>*). The retention of the semantics 'one of the wives in polygamy (in relation to other wives)' in S-H and Alt. reflects the archaic situation, when all women of a particular marriage class were the potential wives of each man of a separate marriage class (punalual family: see Shternberg 129-284 for a more detailed description of a similar system in the Nivkh culture).

# Vedic *mriyáte* and other pseudo-passives: notes on an accent shift<sup>1</sup>

Leonid Kulikov

*Leiden University / Institute of Oriental Studies, Moscow*

## Vedic -ya-presents: introductory remarks

According to the *communis opinio*, Vedic -ya-formations with the accent on the suffix (*kriyáte* 'is made', *dīyáte* 'is given', *hanyáte* 'is killed', etc.) are passives, while forms with the accent on the root (class IV in traditional notation: *jāyate* 'is born', *pádyate* 'falls', *riyate* 'flows', etc.) are not.<sup>2</sup>

There are, however, some exceptions to this distribution, which have forced several scholars to believe that the boundary between passives and non-passives cannot be drawn with accuracy. I quote here only one statement, which is very typical for standard grammars of Vedic: "... der Akzent ist in der älteren Zeit kein unbedingtes Unterscheidungsmerkmal der beiden Präsensbildungen (-yá-passives as opposed to class IV. - LK), da gelegentlich Schwanken herrscht." (Thumb - Hauschild 1959: 333-334)

This opinion seems too pessimistic, however. It will be argued below that the apparent exceptions can be explained if formal and semantic relations between various classes of -ya-presents are better defined.

## Stable vs. fluctuating accentuation

First of all, it is necessary to distinguish between -ya-presents with stable accentuation and those with unstable, or fluctuating, accentuation.

-ya-presents with fluctuating accentuation (*kṣīya-<sup>-te</sup>*/*kṣīyá-<sup>-te</sup>* 'perish', *múcyā-<sup>-te</sup>*/*mucyá-<sup>-te</sup>* 'become free, be released', etc.), generally taken to belong

---

<sup>1</sup> I am much indebted to R.S.P. Beekes and A. Lubotsky for critical remarks to the earlier drafts of this paper.

<sup>2</sup> Semantically, the latter group is more heterogeneous. Intransitives clearly predominate, but a few well-attested transitive -ya-presents belong here (*ásyati* 'throws', *īsyati* 'sends', etc.).

either to *-yá*-passives, or to middle class IV presents, must be treated as a separate group. This small group (less than 20 roots) displays a number of common features: *-yá*-presents are intransitive and mostly denote various kinds of destruction or destructuring. Most of them are opposed to transitive-causative presents with nasal affixes (cf. *kṣināti* 'destroys', *muñcāti* 'releases', etc.). If we look at the distribution of these presents among different texts, we see that there is no free variation in the place of the stress in these formations. More specifically, several texts (Atharvaveda and some Brāhmaṇas) have the accent on the suffix, whereas in the Taittirīya-Saṃhitā this group is usually root-accented (for details, see Kulikov, forthcoming).

As for *-ya*-presents with stable accentuation (i.e. those which always have the accent either on the suffix or on the root), they follow the above-mentioned distribution (passives with the accent on the suffix vs. non-passives with the accent on the root) quite consistently. In particular, it turns out that *-ya*-presents with stable root accentuation (class IV) never show a passive meaning.

Thus, exceptions we have to account for are *-yá*-presents with non-passive meaning. In total, three such presents are found:<sup>3</sup> *mriyáte* 'dies', which is the parade example, mentioned by all grammars, and two more presents, viz. *dhriyáte* 'holds (to), determines' and *driyáte* 'heeds' (cf. Whitney 1896: 277; Macdonell 1910: 333).<sup>4</sup> These presents are attested with middle inflexion only.

It is clear that the meaning of these three *-ya*-presents is not passive, whatever definition of passive we use (for that reason I label them "pseudo-passives"). It would be appropriate to clarify their position within the Vedic verbal system.

### Morphological types and their system-related features

A synchronic system imposes a set of features, such as meaning types, possible syntactic patterns, paradigmatic properties, etc., on its members.

---

<sup>3</sup> I do not discuss here one more non-passive *-yá*-present which might be qualified as exception, *lipyáte* 'stains, sticks'. This present occurs accented only once, in the MS. It can be shown that *lipyáte* should be grouped together with *-ya*-presents with fluctuating accent, i.e. that forms with the accent on the root are only by chance unattested (cf. Kulikov, forthcoming).

<sup>4</sup> *ā-priyáte* mentioned by Whitney among non-passive *-yá*-presents is likely to have been included in this list by mistake. I could not find it in accented texts.

Thus, the affinity of items belonging to the same morphological type is not limited to purely morphological similarity (ablaut grade of the root, suffixation, etc.). The shared features rather form a cluster of properties which goes beyond the morphology, encompassing also paradigmatics, syntax and semantics.

Thus, scrutinizing "non-morphological" features of the three *-yá*-presents in question may be helpful for clarifying their position among verbal formations.

The closest "neighbours" of *-yá*-passives within the system of Vedic present formations are middle *-ya*-presents with root accentuation (class IV). It is therefore plausible to assume that verbs of the type *mriyáte* have more in common with this morphological class than with *-yá*-passives, in spite of their actual accentuation. Thus, before proceeding to the analysis of the type *mriyáte* we have to discuss the semantic and syntactic features of middle *-ya*-presents.

### Middle *-ya*-presents: semantic and syntactic properties

The root-accented *-ya*-presents with middle inflexion can be subdivided into three semantic groups:

(i) Intransitive presents denoting a motion, position or change in body posture: *īya*-<sup>1e</sup> 'move, speed', *ṛjya*-<sup>1e</sup> 'stretch', *pādyā*-<sup>1e</sup> 'move, fall', *rīyā*-<sup>1e</sup> 'flow', *līyā*-<sup>1e</sup> 'adhere, cling' (root *li*<sub>1</sub>, cf. Gotō 1987: 279).

(ii) Transitive presents denoting mental activity: *kāyā*-<sup>1e</sup> 'seek, yearn', *būdhyā*-<sup>1e</sup> 'perceive' (AV +), *mānyā*-<sup>1e</sup> 'think', *mṛśyā*-<sup>1e</sup> 'neglect, forget'.

(iii) Only two of the remaining middle *-ya*-presents are attested in the RV, viz. *jāyā*-<sup>1e</sup> 'be born' and *būdhyā*-<sup>1e</sup> '(a)wake'. Together with *līyā*-<sup>1e</sup> 'dissolve' (Kh., AV +; root *li*<sub>2</sub>, cf. Gotō, *ibid.*), they can be grouped together under the label "intransitive presents denoting change of state, transition from one state to another".

Other *-ya*-presents (all intransitive) appear in later Vedic texts and do not form a well-defined semantic class: *dīpyā*-<sup>1e</sup> 'shine', *rādhya*-<sup>1e</sup> 'succeed', *vāśyā*-<sup>1e</sup> 'bellow'.

Despite the small range of groups (i-iii), their relevance within the verbal system should not be underestimated. These types determine which meanings are productive (and, hence, "morphologically influential") in the class of middle *-ya*-presents, and which are not. In particular, the relevance of type (ii) may account for the secondary and more recent usage of *būdhyā*-<sup>1e</sup>, originally (in the RV) attested only as intransitive 'awaken': after the RV, when class I present *bódhati* 'perceives' dies out, *būdhyā*-<sup>1e</sup> takes



over this usage and meaning ('perceive') and appears in transitive constructions,<sup>5</sup> thus being adjusted to presents like *mānya-<sup>te</sup>*, *mṛṣya-<sup>te</sup>*, etc.

Similarly, *līya-<sup>te</sup>* 'adhere, cling', which appears from the Brāhmaṇa period onward and replaces the older present *lāyate* 'id.' (cf. Gotō, op.cit.), may have been formed under the influence of type (i) (motion, position, etc.).

Taking into account the above-discussed features, we may now turn to the question whether *mriyāte* and the other pseudo-passives can be grouped together with middle *-ya*-presents, at least from the point of view of their semantic and syntactic properties.

#### *driyāte* 'heeds, regards' Br. +

This verb is attested from the Brāhmaṇas onward, mostly with the preverb *ā*. An accented occurrence is found only once, in the ŚB:

*sá yó haitám mṛtyúm ānatimucyāthāmúm lokám éti yáthā haivāsmīmīl  
loké ná samyátam ādriyāte yadā yádaivá kāmāyaté 'tha mārāyaty  
evám u haivāmúsmīmīl loké púnah-punar eva prāmārayati*  
(ŚB 2.3.3.8)

'And whosoever goes to yonder world not having escaped that Death, him he causes to die again and again in yonder world, even as, in this world, one regards not him that is fettered, but puts him to death whenever one wishes.' (Eggeling)

Obviously, *driyāte*, due to its semantics and transitive syntax, corresponds to middle *-ya*-presents (mental activities).

#### *dhriyāte* 'holds (to); decides, determines' RV +

The meaning attested in earlier texts belongs to the semantic domain of change of position and/or body posture, cf.:

*durgé caná dhriyate víśva ā purú  
janó yó asya táviṣīm ācukrudhat* (RV 5.34.7)  
'Even a whole tribe which has made angry his (Indra's) power cannot hold in a fortress'

<sup>5</sup> Cf. Gotō 1987: 219, fn. 459.

The meaning 'determine' appears in Late Vedic (Br. +) and is even further from the passive domain. Cf.:

*svāhāgnīm pávamānam iti yādi pávamānāya dhriyérant*  
*svāhāgnīm indumantam iti yādy agnāya indumate dhriyéran*  
 (ŚB 2.2.3.20)

'[Then he says]: <...> "Svāhā Agni Pavamāna!" - if they **decide** to [offer to] Agni Pavamāna; "Svāhā Agni Indumat!" - if they **decide** to [offer to] Agni Indumat'

*yād vā etē 'mūrhy ādhriyanta tād evāpy adyā kurvanti* (ŚB 14.4.3.34)  
 'What they **determined** then, that they do today also'

This secondary meaning also belongs to the semantic domain of a subclass of middle *-ya*-presents (class ii: mental activities). Thus, not only the original usage of *dhriyáte* can be grouped together with middle *-ya*-presents, but also the later semantic developments are still in accordance with the range of meanings attested in this class.

*mriyáte* 'dies' RV +

*mriyáte* never appears as passive (cf. Jamison 1983: 150, fn.92) and can be easily grouped together with verbs of subclass (iii), which describe transitions from one state to another, cf. esp. *jāyate* 'is born'. Accented forms are attested from the AV onward, cf.:

*striyā yān mriyāte pātih* (AV 12.2.39)  
 '... if a woman's husband **dies**'

There is yet another feature which links *mriyáte* with class IV. The passive meaning is expressed by *-yá*-presents and by middle forms outside the system of the present (cf. *dhriyáte* 'is put' // med.perf. *dadhé* 'has been put', etc.), but never by active forms. In contrast, active forms can be employed in the same usage as corresponding middle *-ya*-presents (non-passive intransitives), cf. *pādyate* 'falls' // act.perf. *papāda* 'has fallen'. This is also the case with *mriyáte*: we find active non-present forms employed in the same usage and with the same meaning ('die') as *mriyáte*, cf.:

*só cín nú ná marāti nó vayām marāma* (RV 1.191.10 = 1.191.11)  
 'Verily he will not die, and we will die neither'

### The type *mriyáte*: a diachronic explanation

The above-discussed semantic features of *driyáte*, *dhriyáte* and *mriyáte* clearly point to their original membership in class IV, despite their suffix accentuation, as is shown in the table below:

-yá-presents	<p style="text-align: center;">passives (<i>kriyáte</i>, <i>dīyáte</i>, <i>hanyáte</i>, etc.)</p>		
	<i>dhriyáte</i> 'holds (to)'	<i>driyáte</i> , <i>dhriyáte</i> 'determines'	<i>mriyáte</i>
middle -ya-presents	<u>motion, position</u> ( <i>pádyate</i> , <i>rīyate</i> , etc.)	<u>mental activity</u> ( <i>mányate</i> , <i>mīsyate</i> , etc.)	<u>change of state</u> ( <i>jáyate</i> , <i>búdhate</i> )
	i	ii	iii

A key to the problem may be a striking morphophonological peculiarity shared by all these presents: they are derived from  $C_r$  roots and, together with -yá-passives of the same structure (*kriyáte* 'is made', *bhriyáte* 'is brought' etc.), represent a specific development of  $r$ . There must be then, I suppose, a phonological reason for the merger of both types *kriyáte* (< \**kryáte*) and *mriyáte* (< \**mryate*). Since the sequence - $\dot{r}y$ - is unattested, we can speculate that the phonetically regular reflex of \* $C_r\dot{r}iV$ - was such that it disturbed the transparency of the formation (for instance, \**mūryate*, \**mūryate* ??). The only way to preserve the transparency of the form was to introduce the accent on the suffix: \* $C_r\dot{r}yā-$  → *Criyá-*. Here the type *kriyáte* (where -*ri-* goes back to an accentless - $\dot{r}$ - before - $\dot{i}$ -) may have served as a model.

Due to this shift, presents like *mriyáte*, which have originally belonged

to middle *-ya*-presents, formally fell together with *-yā*-passives.<sup>6</sup>

*sriyate* 'runs, stretches' KS<sup>1</sup>

One more present can be appended to the group of pseudo-passives, viz. *sriyate*, in spite of the fact that this form is found in an unaccentuated part of the Kāthaka-Saṃhitā:

*so 'napobdho vīryāya prasriyate* (KS 11.4:148.9)  
'He, unbound, stretches to the heroic power' (cf. Narten 1969: 92)

It is clear that this verb has no passive meaning and must be grouped together with middle *-ya*-presents of subclass (i) (motion etc.), cf. esp. the synonymous *ṛjyate* 'stretches'.

On the other hand, although accented occurrences are not attested, the underlying accentuation cannot be anything but *\*sriyāte*, in virtue of the above-formulated accentual rule.<sup>7</sup>

### Conclusions

It has been argued that verbs of class *mriyāte* display a number of features which link them to the middle *-ya*-presents. Despite the "passive" accent of *mriyāte*, this present is never found with passive meaning. Moreover, the meaning 'die' is expressed by active forms outside the system of the present, which is a feature typical of class IV verbs. Finally, the semantic development of *dhriyāte* ('determines') in late Vedic texts complies with constraints imposed on possible meaning types of middle *-ya*-presents. This means that verbs of the type *mriyāte* were still regarded as "surface substitutes" for middle *-ya*-presents, rather than *-yā*-passives proper.

---

<sup>6</sup> It is worth mentioning that this rule, albeit never explicitly formulated in the literature, has been tacitly adopted by some scholars, cf. the following remark by Kellens (1984: 121, note (8)): "Le sens ne permet pas de considérer *mriyā-* comme le passif de *māra-*: l'accent suffixal paraît donc secondaire".

<sup>7</sup> Narten (ibid.) labels this form as "Passiv-Präsens", despite the lack of accent and non-passive meaning, thus, most likely, tacitly relying upon the same assumption.

## Abbreviations

AV - Atharvaveda, Br. - Brāhmaṇas, Kh. - Khilāni, KS - Kāthaka-Saṃhitā,  
MS - Maitrāyaṇīya-Saṃhitā, RV - Ṛgveda, ŚB - Śatapatha-Brahmana.

## References

- Gotō, Toshifumi  
1987 *Die "I. Präsensklasse" im Vedischen: Untersuchung der vollstufigen thematischen Wurzelpresentia.* (= Österreichische Akademie der Wissenschaften. Philos.-Historische Klasse. Sitzungsberichte, 489). Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Jamison, Stephanie W.  
1983 *Function and form in the -āya-formations of the Rig Veda and Atharva Veda.* Göttingen: Vandenhoeck und Ruprecht.
- Kellens, Jean  
1984 *Le verbe avestique.* Wiesbaden: Reichert.
- Kulikov, Leonid  
Forthcoming "Vedic -ya-presents: semantics and the place of the stress." *Akten der 10. Fachtagung der Indogermanischen Gesellschaft.* Innsbruck.
- Macdonell, Arthur A.  
1910 *Vedic grammar.* Strassburg: Trübner.
- Narten, Johanna  
1969 "Ai. *sr* in synchronischer und diachronischer Sicht." *Münchener Studien zur Sprachwissenschaft* 26: 77-103. (= *Kleine Schriften*. Bd. I: 125-143. Wiesbaden: Reichert, 1995)
- Thumb, Albert, and Richard Hauschild  
1959 *Handbuch des Sanskrit. II. Teil: Formenlehre.* Heidelberg: Winter.
- Whitney, William D.  
1896 *A Sanskrit grammar.* Leipzig: Breitkopf & Härtel.

# The Polygenesis of Western Yiddish--and the Monogenesis of Yiddish<sup>1</sup>

Alexis Manaster Ramer  
Wayne State University

With the help of Meyer Wolf, *Natrficial LLC*<sup>2</sup>

1. Mono- vs. polygenesis is an issue that arises not only when we consider the origin of the totality of the world's languages, or of such large sets of languages as the Altaic or Nostratic ones. For example, thought it may come as a surprise to nonspecialists, the single origin of all Yiddish dialects, i.e., their descent from a single Proto-Yiddish ancestral form, is not universally accepted (and, indeed, that in the history of modern Yiddish linguistics, this has been a distinctly minority view).<sup>3</sup> Most scholars (e.g., Landau 1895, Bin-Nun 1973:83 [written in 1935], Birnbaum 1954, M. Weinreich 1973 [1980:726-731] and passim, Marchand 1987, Wexler 1991:28, etc.) explicitly or implicitly assume that different Yiddish dialects may be underlain by different German dialects or different mixtures of various German dialects (together with different admixtures of non-German elements), and only a very few explicitly argue unambiguously for the monogenesis of Yiddish (e.g., Katz 1983:1018; but cf. the qualifications in Katz 1987 and King 1987).

On the other hand, one thing that does seem to be agreed on by all specialists going back as far as Landau (1895:47), arguably the father of modern Yiddish linguistics, is that Yiddish dialects fall into two basic divisions, Western Yiddish and Eastern Yiddish. Western Yiddish, on this view, is the group of dialects spoken (or formerly spoken) in the Netherlands, Alsace, Switzerland, Northern Italy, Germany, Austria, Bohemia, and western Slovakia, together with adjacent parts of Hungary and a small island in southern Poland, and reflected in

---

<sup>1</sup> Many thanks to Marvin Herzog and Meyer Wolf for endless inspiration and encouragement for my work on this and other topics in Yiddish.

<sup>2</sup> I owe a distinct debt of gratitude to Meyer Wolf for selflessly helping put this paper together from my nearly decade-old notes. This special circumstance makes it particularly important to emphasize that the views expressed here, and any errors, are my sole responsibility.

<sup>3</sup> I would like to emphasize, once and for all, that in this paper I am concerned solely with the question of the origin and relationships of the spoken Yiddish dialects attested over the last two or three centuries (and such written language as was/is based on any of these dialects). I have nothing whatever to say about the language of earlier texts composed in a German-based language written in Hebrew characters (texts which many scholars have argued about labeling as 'Yiddish' or as 'Judeo-German'). 'Proto-Yiddish' for me means simply the putative source of all modern Yiddish dialects, assuming that all these dialects do in fact come from a single source. It would not be at all paradoxical if some of the early Yiddish/Judeo-German texts turned out to have come into existence earlier than the date of Proto-Yiddish. However, I do not at present wish to make any claims about what that date was.

the chronologically earlier of the two Yiddish literary languages (the now long-obsolete "Written Language A", in the terminology of M. Weinreich).<sup>4</sup> Eastern Yiddish, on the same view, comprises the remaining Yiddish dialects, spoken (or formerly spoken) in Lithuania, Courland, Belarus, Ukraine, Poland, eastern Slovakia, most of Hungary, Romania, and the Holy Land and forming the basis of the second Yiddish literary language (M. Weinreich's "Written Language B"), including its current variant, Standard Yiddish.<sup>5</sup>

The claim implicit in this classification (and sometimes made explicitly) is that the division into Western Yiddish and Eastern Yiddish is the oldest one in the family tree of Yiddish dialects, an "Urverteilung" in the words of Katz (1983:1024). In these, quite generally accepted terms, then, the question of the unity of Yiddish becomes essentially that of whether Western and Eastern Yiddish come from the same source. Those who appear to (or really do) accept the unity of Yiddish thus hold that the way to arrive at Proto-Yiddish is to first reconstruct a Proto-Eastern Yiddish and a Proto-Western Yiddish and then to compare the two (e.g., King 1987).<sup>6</sup> On the other hand, those who argue that Yiddish dialects do not come from a single Proto-Yiddish see the distinction between Western Yiddish and Eastern Yiddish as even more basic, since it is these two entities to which they normally assign disparate origins, each one involving a different constellation of medieval German dialects and non-German elements (e.g., Bin-Nun 1973, Birnbaum 1954, Marchand 1987, Wexler 1991:28, etc.). Only rarely, as in Marchand's work, is the unity of Eastern Yiddish itself questioned, but, as far as the unity of Western Yiddish is concerned, I have seen no dissent from this in the literature.

Yet, in this paper, I argue that the oldest dialect division within Yiddish cannot have been that between Western Yiddish and Eastern Yiddish. At the risk of adding to the approximate but suggestive fluvial terminology which has become popular in Yiddish linguistics in recent decades (in the course of discussions over a Rhenish as opposed to a Danubian origin of Yiddish), I deny that the oldest split within Yiddish was that running along (or slightly to the east of) the Oder, which is the boundary between Western and Eastern Yiddish. Instead, I advance some reasons for supposing that the earliest split of Yiddish

---

<sup>4</sup> Landau was apparently unaware of the survival into modern times of Western Yiddish dialects in Alsace, Switzerland, or Holland. What he denied was a direct connection between those Yiddish dialects he knew (Eastern Yiddish and the Western Yiddish of Bohemia, Slovakia, and Hungary) with the Written Language A and with whatever Yiddish dialects might still be spoken in Germany.

<sup>5</sup> Occasionally, as in Prilutski (1920:79), the "Central Yiddish" of Poland is excluded from Eastern Yiddish and taken to be a third branch of Yiddish, but this is a view which appears to have no adherents today--and in any case this point would not bear on the question of whether Western Yiddish is a valid geolinguistic concept, which will turn out to be the main issue occupying us here.

<sup>6</sup> A similar view is implicit in any work which assumes that the way to demonstrate that any given feature of Yiddish is Proto-Yiddish is to document its existence in Eastern Yiddish and somewhere in Western Yiddish.

into dialects occurred either along or west (perhaps quite far west) of the Elbe. Consequently Proto-Yiddish originally divided into a 'Westerly Yiddish' and an 'Easterly Yiddish' (the latter encompassing Eastern Yiddish as well as the easterly dialects of what is generally conceived of as Western Yiddish), rather than into Western and Eastern Yiddish, as usually assumed.

If this hypothesis is correct, then 'Western Yiddish' is not even a valid concept in the historical dialectology of Yiddish (because some Western Yiddish dialects are more closely related to Eastern Yiddish than they are to the rest of Western Yiddish), and Eastern Yiddish, instead of being a direct descendant of Proto-Yiddish, becomes a mere offshoot of some intermediate proto-dialect which also gave rise to much of what is traditionally considered to belong to Western Yiddish. All this has an unexpected consequence for the question of the original unity of Yiddish. For, if all investigators accept--and if they are right to accept--that all so-called Western Yiddish dialects have a common origin, and, if at the same time, some of these dialects (the more easterly ones) actually are more closely related to Eastern Yiddish than they are to other (more westerly) Western Yiddish ones, then it would have to follow that all of Yiddish does in fact come from a single source. This I believe to be the correct result, for which there is massive evidence of various other kinds as well (evidence which is briefly sketched below and which we hope to be able to present in detail in a monographic treatment some time in the future).

2. To turn to specifics, when it comes to defining Western Yiddish, there has been, ever since Landau and Wachstein (1911:xli), what amounts to a single criterion: the realization of the Proto-Yiddish<sup>7</sup> vowel E<sub>4</sub> (corresponding to Middle High German /ei/ and /öu/) as well as that of the Proto-Yiddish vowel O<sub>4</sub> (corresponding to Middle High German /ou/) as /a:/ in Western Yiddish (e.g., Beranek 1961:281, U. Weinreich 1964:252, Katz 1983, etc.), although many authors for reasons which are completely unclear to me mention only E<sub>4</sub> (e.g., Prilutski 1920:79, M. Weinreich 1958b, Garvin 1965:94, etc.). There are some other features which are frequently mentioned as being characteristic of Western Yiddish (such as traces of the preterite and certain lexical items), but I know of no such features (be they phonological, morphological, or lexical) which are shared innovations of all of Western Yiddish to the exclusion of Eastern Yiddish. Rather such features as the absence of the preterite involve Eastern

<sup>7</sup> The generally accepted system of Yiddish proto-vowels is that of M. Weinreich (1973), to which Katz (1983) has proposed some crucial revisions. To be sure, M. Weinreich himself did not believe in Proto-Yiddish (a concept he had entertained, but with considerable reservations, in M. Weinreich 1940). Thus, even though he did posit a system of what he himself called 'kadmen-vokaln [which corresponds precisely to English 'proto-vowels', but got rendered as 'Early Vowels' in the posthumous (1980) translation], he held these to be essentially an "ideal" scheme of correspondences between the vowel systems of the different attested dialects, not necessarily the vowel system of a real proto-language from which these dialects are derived.



Yiddish innovations. This is quite easy to see because the preterite is clearly of ultimately Proto-Germanic origin (and survived, if not in full, then in trace form, in Proto-Yiddish), so the innovation here is the loss of it in Eastern Yiddish—not the retention in Western Yiddish. It is, as Leskien taught us more than a century ago, only shared innovations which demonstrate linguistic unity,<sup>8</sup> and so the fate of the preterite does **not** argue for the internal unity of Western Yiddish.

The same is true of the many lexical isoglosses mentioned in the literature (e.g., U. Weinreich 1961: 64, Katz 1983:1025) as separating Western Yiddish from Eastern Yiddish, such as Western Yiddish *barkhes/berkhes* 'challah', *ete* 'father', *fra:le* 'grandmother', *harle* 'grandfather', *meme* 'mother', *o:rn* 'to pray', *minikh* 'neither meat nor dairy', etc. (as opposed to Eastern Yiddish *khale*, *tate*, *bobe*, *zeyde*, *mame*, *davenen*, *parev(e)*, etc., respectively). In each of these cases (just as in the case of the preterite), we may assume that the Eastern Yiddish forms represent innovations, while the Western Yiddish words simply continue Proto-Yiddish lexical items. If, for example, Proto-Yiddish had *\*frE<sub>4</sub>/O<sub>4</sub>le*,<sup>9</sup> rather than *\*bA<sub>2</sub>/A<sub>3</sub>be*, for 'grandmother' (and similarly in all the other cases), then there is no *prima facie* case for assuming any kind of unity among those dialects which still have *fra:le*. Moreover, the specifics of these cases are such that what could have been a purely a priori argument based on Leskien's teaching becomes an empirical one as well. For the fact is that in almost every case the data strongly point to Eastern Yiddish having replaced Proto-Yiddish words which were retained in Western Yiddish. This is because the Eastern Yiddish forms are typically Slavic or Hebrew, whereas Western Yiddish has Germanic or Romance ones.<sup>10</sup> It is more likely that a Germanic or Romance form would have been replaced by a Slavic or Hebrew one than the other way around, given what we understand of the dynamics of the evolution of Yiddish, and hence in almost every case it is immediately clear that the innovation is an Eastern Yiddish one (and in the cases where this is not obvious, it is at least possible).

In addition, there is another reason why these lexical isoglosses do not argue for the supposed Eastern Yiddish/Western Yiddish dichotomy. Typically, these lexical isoglosses do not coincide with the phonological isogloss between the dialects where E<sub>4</sub> and O<sub>4</sub> changed to /a:/ and those where this change did not take place. Lowenstein (1969), based on fieldwork done in the 1960s, identifies

<sup>8</sup> Actually, as I propose to discuss elsewhere, Leskien's position, which has become part of the canon of comparative linguistics, is not entirely correct. It seems clear, on simple probabilistic grounds, that the accumulation of a sufficiently large number of identical shared retentions can also serve as evidence of kinship.

<sup>9</sup> In light of the discussion below, the Alsatian form *fra:le* implies that Proto-Yiddish had *\*frO<sub>4</sub>le* (corresponding to a MHG *vrouwelîn*) and not, as seems to be generally assumed, *\*frE<sub>4</sub>le* (MHG *vröuwelîn*). This issue requires further investigation.

<sup>10</sup> Occasionally, both Eastern Yiddish and Western Yiddish have Hebrew forms but different ones, and some of the words (especially perhaps in Eastern Yiddish) have no generally accepted etymology, e.g., *davenen*, *parev(e)*, etc.

only a single lexical isogloss which coincides with the phonological boundary between Western Yiddish and Eastern Yiddish, namely, the *barkhes/khale* line, and this can therefore be regarded as a likely coincidence (or, perhaps, a late development, irrelevant to the early prehistory of Yiddish). The isoglosses for the other lexical items which are supposed to separate Western Yiddish from Eastern Yiddish fall either further east (i.e., within Eastern Yiddish territory itself) or, more commonly, much further west (i.e., within Western Yiddish territory). Indeed, U. Weinreich (1961:64), who views the division Eastern Yiddish and Western Yiddish as reflective of a fundamental cultural and ritual, and not just linguistic, cleavage within Ashkenazic Jewry, says:

Tsrur ha-isoglosim ha-mavdil beyn mizrakh le-ma'arav  
nimshakh--mesha'erim anu--le-orekh nahar ha-Elba derekh  
Germanya ve-Beyhem... [The bundle of isoglosses which  
divides East from West stretches --we suppose --along the Elbe  
River through Germany and Bohemia...]<sup>11</sup>

Now, by common consent, the western boundary of Eastern Yiddish and the eastern boundary of Western Yiddish as defined by the  $E_4/O_4 > /a:/$  isogloss lies well east of the Oder. It is no accident that older works sometimes roughly identify it with the pre-1939 western boundary of the Polish Republic (e.g., M. Weinreich 1940), or that the westernmost dialect of Eastern Yiddish (Central Yiddish) is commonly called "Polish Yiddish". Yet the same boundary as defined on the basis of lexical isoglosses cuts across Germany, along the Elbe, and so lies much further west than the boundary based on the phonological isogloss  $E_4/O_4 > /a:/$ , which is usually taken as criterial for Western Yiddish. Also, there is an enclave of  $E_4/O_4 > /a:/$  well within the body of Central Yiddish, in the Sosnowiec/Będzin area of southern Poland (Prilutski 1920), but there is no evidence that this enclave shares the lexical properties of Western Yiddish. Therefore, Beranek (1961:281) was, if anything, understating the case when he said, referring to the change of  $E_4$  and  $O_4$  into  $/a:/$ :

Von den im Grunde nur wenigen Kriterien, durch welche sich  
das Westjiddische in seiner Gesamtheit vom Ostjiddischen  
abhebt, reicht auch nur ein einziges mit seinen Wurzeln bis  
ins Urjiddische zurück. [Of the basically few criteria which set  
Western Yiddish in its entirety stands apart from Eastern  
Yiddish, only one has its roots in Proto-Yiddish.]

<sup>11</sup> Throughout, all translations as well as the transliterations of (Israeli) Hebrew (using an admittedly somewhat idiosyncratic system, which, however, seems much easier to use than other systems and, what is most important here, is consistent with the conventional transliteration of Yiddish) are mine.

However, Beranek's statement is only correct in so far as it focuses on the fact that other isoglosses do not coincide even approximately with this one. There is, on the other hand, absolutely no reason to agree with him that this isogloss "reicht mit seinen Wurzeln ins Urjiddische zurück". As I propose to show, this cannot be the case, because there are other isoglosses which must antedate the merger of  $E_4$  and  $O_4$  into /a:/ and which divide Western Yiddish into two (or more) parts along boundaries which must accordingly must be older than the isogloss in question. In other words, the change of  $E_4$  and  $O_4$  to /a:/ must be a relatively recent development, which took place after the Western Yiddish dialects had already subdivided in some fashion. And this in turn means that this sound change cannot be used to define Western Yiddish as a historically valid dialectological unit.

3. In fact, Beranek himself goes on in the same paragraph to refer obliquely to the fact that the development of  $E_4$  and  $O_4$  to /a:/ does **not** represent the earliest realization of these two vowels in Western Yiddish generally, when he says, referring to the /a:/ realization of Proto-Yiddish  $E_4$  and  $O_4$  (what he calls "die Lautung /a:/):

...doch hat ihr [sc. der Lautung /a:/] bei der Ausbildung des westjiddischen Lautstandes--außer ihrer phonetischen Einfachheit und Klarheit--sicherlich in erster Linie der Umstand zum Siege über andere konkurrierende Lautungen verholfen, daß sie auch der Sprache der Judenmetropole Frankfurt am Main sowie der alten Stadt- und Verkehrsmundart der Sudetenländer und Österreichs angehörte. [...but, in the development of the Western Yiddish sound system, what contributed to the victory of this [sc. the /a:/ pronunciation] over other, competing realizations--apart from its phonetic simplicity and clarity--was surely above all the fact that it occurred in the language of the Jewish metropolis of Frankfurt am Main as well as in the old urban and trade dialect of the Sudeten and Austria.]

Here Beranek is clearly assuming that the change of  $E_4$  and  $O_4$  to /a:/ originated in a small part of Western Yiddish and then spread over the rest of Western Yiddish, replacing earlier pronunciations there. The same position was in effect taken by M. Weinreich (1958b) in his classic work on Western Yiddish, where he seems to trace it to the 15<sup>th</sup>-century speech of Jews in the major urban centers of Germany and Austria, whence it would have spread across all of Western Yiddish.

However, to even say what I just did makes no real sense in diachronic terms, because there was no "Western Yiddish" for the /a:/ pronunciation to spread across. It is precisely this pronunciation which is criterial for Western

Yiddish, and hence there was no such thing as Western Yiddish before this pronunciation became entrenched in those areas in which we subsequently find it. Western Yiddish, to the extent that we recognize this as a valid concept at all, came into being precisely in virtue of the fact that the sound change in question spread precisely to those areas where it did and to no others. If earlier there were other competing pronunciations, then that would mean that there must have been divisions within what now came to be Western Yiddish, and hence the dichotomy between Eastern and Western Yiddish is far from the oldest dialect division within Yiddish. This is indeed clearly the case. Beranek (1965:8, 10) reports scattered traces of relic forms with /e:/ for  $E_4$  and similar traces of relic forms with /o:/ for  $O_4$  in Yiddish dialects of the Rhineland, an area which in general, like all of Western Yiddish, has /a:/ in both cases. From the existence of these relic forms, we immediately deduce that Proto-Yiddish could not have first split into Western Yiddish and Eastern Yiddish, since, by the time of the criterial sound change for Western Yiddish, it will already have split at least into two dialects: a conservative variety (in the Rhineland) where the contrast between  $E_4$  and  $O_4$  was preserved and an innovative one (elsewhere, perhaps starting in the Frankfurt area) where the change to /a:/ originated. The fact that this earlier split was almost completely obscured by the subsequent spread of the /a:/ pronunciation does not alter the fact that the earlier split had happened--and that it happened, by definition, before the spread of /a:/, hence before the event which by the common consent of Yiddishists defines the formation of Western Yiddish.<sup>12</sup>

4. Another possible argument against the unity of Western Yiddish has to do with the distribution of the palatal fricative /ç/, the so-called *ich-laut*, in those Western Yiddish dialects that have it (e.g., Switzerland), or equivalently the distribution of /ʃ/ in those dialects where the *ich-laut* has merged with /ʃ/ (e.g., Alsace and Holland). These sounds, as is well known, appear in place of the *ach-laut* /x/ after originally front vowels and diphthongs, crucially including  $E_4$  (= MHG /ei/), even when this has merged with  $O_4$  and turned into /a:/. The existence of forms like *va:ç* or (secondarily) *va:ʃ* 'soft' indicates that the /ç/ must have existed **before** the merger of  $E_4$  and  $O_4$  into /a:/. But if so, then there are only two possible explanations.

<sup>12</sup> It may be useful to point out that there is nothing in the least unusual about a sound change spreading over a territory which had already undergone some dialectal differentiation earlier. Herzog (1969:77-80) in his discussion of Proto-Yiddish  $U_4$  documents at least three distinct developments which involved "diffusion" of a particular realization of this phoneme across parts of Ukraine--a good analogue to the scenario I am positing for the "Western Yiddish" change of  $E_4$  and  $O_4$  to /a:/. The loss of syllable-final /r/ in various dialects of English in England and overseas is another good example: the non-rhotic (alias *r*-less) dialects are not particularly closely related to each other.

Either, as M. Weinreich (1958a) seems to have believed, the earliest forms of Yiddish (in effect, Proto-Yiddish, although he himself rejected both this term and the concept of a proto-language) had the *ich-laut*, which was then somehow systematically replaced by /x/ again in those dialects which lack such a sound (Eastern Yiddish and easterly Western Yiddish), or else the *ich-laut* arose in those dialects which do have it before the merger of E<sub>4</sub> and O<sub>4</sub> into /a:/. If we believe in a Proto-Yiddish /ç/, then nothing follows, but this seems a very questionable--though possible--belief.<sup>13</sup>

On the other hand, if we believe that the /ç/ was an innovation of some westerly dialect(s) of Yiddish (much as it was an innovation in parts of German only, excluding for example Alemannic and Bavarian), then it would have to follow that this innovation involved a split within Yiddish which antedates the spread of /a:/ across all of Western Yiddish. Yiddish would thus have had a westerly dialect with the *ich-laut* and an easterly one without it, some time before the merger of E<sub>4</sub> and O<sub>4</sub> into /a:/ could have taken place at all.

5. Finally, while I have so far assumed, with almost all the earlier authors, that there is indeed a single /a:/ reflex of E<sub>4</sub> in Western Yiddish, the reality is more complex.

First of all, no distinction between front rounded and front unrounded vowels is posited in the generally accepted system of Yiddish protovowels, and therefore the change of E<sub>4</sub> to /a:/ is claimed to apply equally to both the correspondents of MHG /ei/ and those of MHG /ou/, the latter being found in a small class of words such as (Standard Yiddish) *freyd* 'joy', *freyen zikh* 'enjoy oneself', and *hey* 'hay'. However, Guggenheim-Grünberg (1965:152, n. 8) reports that in the two southwesternmost varieties of Western Yiddish, Swiss and Alsatian Yiddish, MHG /ou/ yields, not /a:/, but /ai/--which is also the reflex of MHG î (= Yiddish protovowel I<sub>4</sub>)--as in /fraid/ 'joy', /siç gfrait/

<sup>13</sup> The only argument I can come up with for a Proto-Yiddish /ç/ might be the fact that Eastern Yiddish dialects have preserved traces of a diminutive ending (cognate with German *-chen*) in the form *-khn* /-xn/ or, secondarily (as in Standard Yiddish), *-khl* /xl/, which are used after stems ending in /l/. Since no German dialect appears to have the *ach-laut* /x/ in this ending, it might be argued that Proto-Yiddish also must have had the *ich-laut* /ç/ here. However, this may be a complete accident due to the way two quite independent isoglosses happen to fall out: the (southern) German dialects which, like Easterly Yiddish, have no *ich-laut* seem to lack the *-chen* diminutive suffix altogether. As for the existence of the *ich-laut* in some diminutive forms in the Eastern Yiddish of Kalisz, Poland, this appears to be a secondary development, probably involving borrowing from German, rather than an inheritance from Proto-Yiddish. Such diminutives are also found in some "Western Yiddish" dialects (spoken in Germany), and it is likely that Kalisz Yiddish acquired them from such a dialect, together with at least one other feature it shares with some "Western" dialects: in Hebrew-origin words, initial /s/ (which of course does not occur in the German component) has changed to /ts/ (which is common word-initially in German).

'enjoyed himself,' and /hai/ 'hay'.<sup>14</sup> In contrast, what she calls "general Western Yiddish" has /a:/ in such cases, e.g., /fra:d/ in Hungarian Yiddish (Shpirn 1926:198, Garvin 1965:101) and in what we may call Central German Yiddish (Prilutski 1920:75, 76, etc., with references to primary sources).

Second, Guggenheim-Grünberg (1965: 153, n. 12) calls attention to Swiss Yiddish /tsvai/ 'two', which she contrasts with "general Western Yiddish" /tsva:/. Although she did not point this out, the significant point here is that MHG /ei/ (i.e., Proto-Yiddish  $E_4$ ) has two distinct reflexes in Swiss Yiddish, /a:/ in general but /ai/ in /tsvai/. Unfortunately, we are not told whether the same situation with the word for '2' obtains in Alsatian Yiddish.<sup>15</sup> But the fact that, for example, Hungarian Yiddish (Hutterer 1965:125), Dutch Yiddish (Beem 1959, entries 193, 220, 650, 698, 705, 1041), and Central German Yiddish (Prilutski 1920:73-79, with references to primary sources) all have /tsva:/ indicates that the shift of  $E_4$  (even excluding the cases, discussed above, that correspond to MHG /öu/) did not proceed in the same way in Swiss (and probably Alsatian) Yiddish as in other dialects. Although I lack any additional data even for Swiss Yiddish, it seems possible that the /ai/ reflex is regular in word-final position.

In short, although much remains to be done on this subject, it seems clear that the formula  $E_4 > /a:/$  conceals one or perhaps two minor but crucially important isoglosses which subdivide Western Yiddish and separate Swiss (and Alsatian) Yiddish from other, more northerly or more easterly dialects.<sup>16</sup> Hence,

<sup>14</sup> Also /gai/ 'region', which has no cognate in Standard Yiddish (but compare MHG *göu*).

<sup>15</sup> Weill's (1920-1921) Alsatian Yiddish form *zweiling* 'jeûne de deux jours consécutifs pour obtenir l'aide de Dieu dans un cas désespéré, pour un malade à toute extrémité' is clearly related but unfortunately we cannot tell how this was pronounced because Weill often uses Standard German spelling for his Alsatian Yiddish, esp. in the keywords to his entries. For example, under the misleading keyword *Kleid*, he has an example sentence containing the correct dialect form *Klâder*. Thus *zweiling* could represent either \*/tsva:ling/ or \*/tsvailing/.

<sup>16</sup> Timm (1987:88, n. 8; 208, n.4) notes some of the same facts as discussed here, and in addition raises the topic of umlauted plurals like /ba:m/ (but /baim/ in some southwestern German Yiddish dialects) 'trees', but does not draw any definite conclusions, except to suggest that the change to /a:/ failed to take place prevocally. However, this cannot be generally true in Western Yiddish, since we find *a:r* /a:r/ 'eggs' from \* $E_4$ er in South-Central Hungarian Yiddish (Garvin 1965:100). She also notes some indications from early Western Yiddish texts that there may have been places where the change of  $E_4$  to /a:/ did not occur before velars or palatals. However, in any case, she fails to specify that we are dealing with crucial east/west isoglosses cutting across Western Yiddish and antedating the sound change considered definitional for Western Yiddish. The plural forms like /baim/ may very well represent additional evidence for a Proto-Yiddish  $\ddot{O}_4$  distinct from  $E_4$ , but the problem is rendered nearly hopeless by the paucity of relevant dialect data together with the possibility of complications due to analogy.

once again, all of so-called Western Yiddish could **not** have undergone the shift of E<sub>4</sub> (i.e., MHG /ei/ and /öu/) to /a:/ at the same time or in the same way.<sup>17</sup>

6. I hypothesize, therefore, that there are several different isoglosses cutting across Western Yiddish at various points no further east than the Elbe which are older than--or at least as old as--the very phenomenon which is conventionally used as the criterion for treating Western Yiddish as a single unit and to differentiate it from Eastern Yiddish. The facts I have cited, although all of them require much more detailed investigation,<sup>18</sup> thus seem to close the books on Western Yiddish as a valid unit in a diachronic classification of Yiddish dialects. Hence, no matter how different Eastern Yiddish looks from Western Yiddish, and no matter how sharp the boundaries between Eastern Yiddish and Western Yiddish dialects are in Slovakia and Hungary, where the two kinds of Yiddish meet face to face, we have no basis for assuming that Eastern Yiddish was at its inception any more different from all the "Western Yiddish" dialects than they were (at the time) from each other.

Perhaps an analogy from a field very different in substance but rather similar in form to linguistics will help make clear how this can be. Recent work on the DNA of primates has shown that genetically *Homo sapiens* is more similar to the two species of chimpanzee (the true chimp and the bonobo) than any of these three species are to any other primate. What this means is that, strictly speaking, the commonsensical category of 'ape', i.e., nonhuman primate, has no validity, for there was no common ancestor of all the "apes" who was not also an ancestor of our own species (and likewise the category 'chimpanzee' is invalid if by 'chimpanzee' we mean the two nonhuman species which are most closely related to us, the true chimp and the bonobo). We can only salvage the terms 'ape' and 'chimpanzee' if we use them so as to include humans. In the same way as in biology, appearances can be deceptive in linguistics as well. For, despite all appearances, we can now see that there was no ancestor of Western Yiddish that was not also an ancestor of Eastern Yiddish. Eastern Yiddish is thus no different in diachronic status from the various dialects of Western Yiddish. If we wanted to persist in using the term 'Western Yiddish', we would then have to say that Eastern Yiddish is a dialect of Western Yiddish (much as Jared Diamond gives *Homo sapiens* the title of *The Third Chimpanzee*). But this would be confusing and wasteful, since **all** of Yiddish would now have to be called 'Western Yiddish'. Hence, we should really stop using the term 'Western Yiddish' altogether.

<sup>17</sup> A preliminary investigation of Dutch Yiddish suggests that it agrees in some respects with Swiss and Alsatian dialects, and in others with the more easterly varieties of Yiddish. This, if correct, would imply still at least one more such isogloss subdividing "Western Yiddish".

<sup>18</sup> I hope to present more information on this topic at a future date, including a detailed study arguing that the stressed vowel of [*dos*] *gayes* 'gentiles, non-Jewish populace' is derived from Ö<sub>4</sub>, for this word is related to Swiss and Alsatian Yiddish *gai* and hence to MHG *göu* (Manaster Ramer, to appear a).

As for Eastern Yiddish, we do not at present know just where, when, or how it split off from the rest of Yiddish. Specifically, we do not know just how differentiated Yiddish already was at the time of Eastern Yiddish's separation. Or, in other terms, we do not know how high in the family tree of Yiddish the node labeled Proto-Eastern Yiddish should go. It is entirely possible that it belongs relatively low down, i.e., after any number of other splits had already taken place. But in any case, Eastern Yiddish is a descendant, not of Proto-Yiddish directly, but of some intermediate proto-system which is also the source of just some but not all of the attested Western Yiddish dialects. The apparently sharp delimitation of Eastern Yiddish and Western Yiddish dialects found today in Hungary and Slovakia is quite real, and it tells us that for some significant time in history the two varieties of Yiddish must have been completely out of touch, so that the immediately neighboring Eastern Yiddish and Western Yiddish dialects cannot be at all closely related to each other. Eastern Yiddish must thus be seen as a dialect which separated cleanly from the other Yiddish dialects and undergone autonomous development for a considerable time--but this has to do with a time in history long after the break-up of Proto-Yiddish into dialects. Originally, Eastern Yiddish was nothing but a dialect of easterly "Western Yiddish" (i.e., of Easterly Yiddish, in my terms).

7. We have seen that there is evidence that the boundary dividing Western Yiddish from Eastern Yiddish is not the oldest dividing line within Yiddish. However, as I have already implied, while we do have evidence of various earlier splits along more westerly boundaries, we do not at present possess sufficiently detailed information about the distribution of the various features in Western Yiddish dialects to be able to say just where the earliest division of Yiddish into dialects actually occurred. Still, it may not be premature to advance a hypothesis, with the understanding that this is merely a suggestion whose utility, if any, will be to focus attention on the urgent need for more research in this area.

We have already seen that distribution of the different realizations of  $E_4$  (and its probable rounded counterpart, which we may call  $\ddot{O}_4$  or perhaps  $O_5$ )--in conjunction with the interaction between the change of  $E_4$  to /a:/ and the palatalization of /x/ to /ç/--leads us to posit one or more dividing lines well within the body Western Yiddish, dividing lines which must antedate the change of  $E_4$  (and  $\ddot{O}_4$  and  $O_4$ ) to /a:/, although we do not know where exactly these lines actually fell. However, it is unlikely that these isoglosses (all running well west of the Elbe), for all that they are certainly older than the isogloss separating the supposed Western Yiddish from Eastern Yiddish, are the oldest. Instead, I am more impressed with U. Weinreich's broad generalization that the major bundle of lexical isoglosses dividing East and West ran along the Elbe River. This need not be taken literally, because, as U. Weinreich notes, the parts of Germany transected by these isoglosses are ones where Yiddish dropped out of use particularly early. All that is really at issue here is that a major dividing line within Yiddish runs somewhere **within** what is supposed to be Western



Yiddish territory, considerably to the west of the boundary between Western Yiddish and Eastern Yiddish--but well east of the isoglosses I have been discussing on the basis of what happens to  $E_4$ ,  $O_4$ , and  $\ddot{O}_4$  in the Rhineland and in Alsace and Switzerland.

It is noteworthy that the distribution of various nonlinguistic boundaries between dividing traditional Ashkenazic Jewry, discussed by Lowenstein (ms.), seems to agree very well with my linguistic conclusions. For example, the linguistic concept of 'Southwesternmost Yiddish' (i.e., the Yiddish of Switzerland, Alsace, and SW Germany) seems to correspond closely to the area where the *holekra:sh* ceremony for naming newborn children was practised among Jews. As for the boundary between Easterly and Westerly Yiddish, we find that east and west of the Elbe there were different prayerbooks, divergences in several rituals, differences in liturgical music, and even in food. For example, "[g]efilte fish was virtually unknown west of the Elbe". A folk awareness of this boundary was reflected in a saying once proverbial among Hamburg Jews, "Poland [i.e., Eastern or Easterly Jewry] begins at the Dammtor" [Hamburg's eastern gate].

Going back to linguistics, it also turns out (and has long been observed, notably by Beranek 1961, 1965) that there are a number of major phonological differences between the easterly varieties of Western Yiddish (which in all cases agree with Eastern Yiddish) and the more (south)westerly varieties. The most striking of these Easterly features include the use of /x/ instead of /ç/, the use of initial /f/ and non-initial /p/ corresponding to MHG /pf/, the use of /p/ in place of /b/ in the words *gopl* 'fork', *nopl* 'navel', the rounding (and raising) of the Proto-Yiddish vowel  $A_3$ . To be sure, these isoglosses themselves do not synchronically, at least, coincide all that well (for example, Dutch Yiddish has, or had, the *ich-laut* but not /pf/). Still, it may be that Beranek was right to generalize that the main dividing line was, once again, in the general neighborhood of the Elbe. There is much work that still needs to be done in this area, but it does not seem an unreasonable working hypothesis that Yiddish first split into two major dialect areas, a westerly one and an easterly one, with the boundary between them running along or somewhat west of the Elbe, that this happened long before there was an Eastern Yiddish (and, of course, before the rise and spread of the /a:/ pronunciation which defines Western Yiddish), and that Eastern Yiddish together with the easterly varieties of Western Yiddish are offshoots of the 'Easterly Yiddish' proto-dialect, whereas the varieties of Western Yiddish spoken further west (in Holland, (south)western Germany, Alsace, and Switzerland) descend from the 'Westerly Yiddish' proto-dialect.

As noted above, the further division of Westerly Yiddish into Southwesternmost Yiddish and the rest of Yiddish is older than the spread of the /a:/ pronunciation characteristic of Western Yiddish (and so are the isoglosses dividing the areas where the /a:/ pronunciation originated from those, such as the Rhineland, where it was a late-comer), but for the moment it seems reasonable to posit (though this is just a surmise) that the isoglosses separating Southwesternmost Yiddish from the rest of Yiddish (as well as those separating

the dialects of the Rhineland from the rest of Western Yiddish, and so on) are more recent than the division between Westerly and Easterly Yiddish. Considerable effort will be needed to settle these issues, but for the moment the sheer weight of the combined lexical and phonological isoglosses seems to point to the Elbe as a plausible candidate for the approximate (or at least symbolic) earliest dividing line between Easterly and Westerly Yiddish. It could turn out that it is the other way around, of course, that is, that one or more of the more (south)westerly isoglosses are older than the ones closer to the Elbe. The issue is factual, and will surely be settled by further work.

8. There is, in short, every indication that Western Yiddish is a spurious construct, and that the earliest splits within Yiddish were that between Westerly Yiddish and Easterly Yiddish (the latter including what would later become Eastern Yiddish, but also the Western Yiddish dialects of Hungary, Slovakia, and eastern parts of Germany) and those between various subgroups of Westerly Yiddish dialects (including but not restricted to those which set off Southwesternmost Yiddish and those which set off Rhineland Yiddish). As I observed at the outset, all this has an unexpected but comforting consequence for my view of the mono- vs. polygenesis of Yiddish as a whole. As noted, I believe that many if not all of those who deny the unity of Yiddish (and indeed those who have hitherto supported it, too) have apparently accepted the unity of Western Yiddish. Yet we now see that the dichotomy between Western Yiddish and Eastern Yiddish is a recent one and hence of no import for the comparative picture.

The logic of the situation before us is simple. If all the dialects belonging to (the spurious category of) Western Yiddish are admitted to come from a single source, that source cannot be a hypothetical Proto-Western Yiddish, because there could not have been such a thing as a Proto-Western Yiddish given the facts described in this paper. Instead, these facts indicate that some of the Western Yiddish dialects derive from the same proto-dialect (Proto-Easterly Yiddish) as does Eastern Yiddish, whereas other Western Yiddish dialects do not. This means that the most recent proto-system from which all of Western Yiddish dialects could derive would be Proto-Yiddish itself, but that of course can only be the case if there **was** such a thing as Proto-Yiddish. The critics of this latter concept will have to either come round to accepting it--or else rethink their views on the matter of Western Yiddish. We have thus scored a significant debating point for the monogenesis of Yiddish.

Of course, ultimately one wants more than debating points. The proof of the unity of all Yiddish dialects cannot be presented in full here, for lack of space (indeed, it seems to call for a monographic treatment). All I can do is point out that I seek to act on the fundamental idea (which was anticipated in large measure by Katz 1983, 1987) that the crucial thing is to identify features which are so distributed across Yiddish dialects that they would be reconstructed for Proto-Yiddish (if one were to posit a Proto-Yiddish) and which make for a picture of Proto-Yiddish which clearly sets the latter apart from any other

language or dialect (in particular, from any German dialect). To be sure, I do not find entirely satisfactory such arguments for the unity of Yiddish as were given by Katz (1983, and with some qualifications, 1987) himself. However, I have myself collected (building on earlier work, of course) a very large number of phonological, morphological, lexical, and even phraseological features shared by Eastern Yiddish with even the westernmost varieties of Yiddish (those of Switzerland, Alsace, (south)western Germany, and Holland) and presumably universal throughout Yiddish (if not at present, then at earlier times)--features which can therefore with virtual certainty be posited for Proto-Yiddish and which by their very existence make the reality of Proto-Yiddish itself also virtually certain.

This forces us to briefly discuss the question of what kinds of linguistic phenomena we should posit for Proto-Yiddish. While this is really a complex topic, for the present some basic observations will have to do. The principal one is that, if there was a Proto-Yiddish, this cannot be validly reconstructed by comparing Proto-Eastern Yiddish and a supposed Proto-Western Yiddish, for the latter cannot have existed at all. If it were true that Yiddish first split into Eastern Yiddish and Western Yiddish, and only then did each of these subdivide into subdialects, it would be right to base our conclusions about the origin and development Yiddish on those facts that can be corroborated by Eastern Yiddish and Western Yiddish evidence. However, if the division into Eastern Yiddish and Western Yiddish was not the primary one, then this methodological principle is too weak **and** too strong. On the one hand, if some Eastern Yiddish feature is shared by the easterly parts of Western Yiddish, but not with all of Western Yiddish, then we may no longer safely assume that we are dealing with a Proto-Yiddish trait. It could be a Proto-Easterly Yiddish one. On the other hand, if some feature is absent from Eastern Yiddish but found in both easterly and westerly Western Yiddish, e.g., in Bohemia and Switzerland, Hungary and Alsace, Slovakia and Holland, or the like, then it is quite likely that such a feature is Proto-Yiddish (a number of the typically "Western" lexical items certainly fit in this category). In any event, it will not do to use the Yiddish of Hungary, Bohemia, or eastern Germany to corroborate features of Eastern Yiddish which we wish extrapolate to Proto-Yiddish. We must look further west, and in particular to the dialects of western Germany, Alsace, Switzerland, and Holland, for such evidence instead, and it may indeed be safest to base our reconstruction in the first instance on those phenomena which recur in Eastern (especially Easternmost) and Westerly (especially Southwesternmost) Yiddish.

The features of Proto-Yiddish which can almost immediately be posited in this way include but are not restricted to:

- (i) a highly distinctive set of Romance lexical items, including, among others, *tscholent* 'a baked dish of meat, potatoes, and legumes served on the Sabbath, kept warm from the day before in view of the prohibition against cooking on the

Sabbath',<sup>19</sup> *leyenen* 'to read', *bentshn* 'to bless', etc.;

(ii) an if any anything even more distinctive set of Slavic lexical items comprising *khotsh(e)* 'although', *nebekh* 'poor thing!', *koyletsh* '[c]hallah' (or, depending on dialect, one of a number of other kinds of pastry),<sup>20</sup> *khapn* 'to grab' (if it really is Slavic, as seems all but certain<sup>21</sup>), and perhaps no others;<sup>22</sup>

(iii) a large set of additional characteristically Yiddish vocabulary whose etymologies and/or particular semantic, morphological, or phonological developments are specifically Yiddish,<sup>23</sup> such as *shmeykhlen* 'to smile' and *grayz*

<sup>19</sup> This definition, taken from U. Weinreich's (1968) dictionary of modern Standard Yiddish, probably is not accurate for Proto-Yiddish, since potatoes are a new world vegetable. It should be noted that, as far as possible (that is, as long as they are attested in Standard Yiddish), I cite all Yiddish words and phrases according to this dictionary, and that accordingly the forms and/or meanings given are usually not quite what we would posit for Proto-Yiddish (or what we find in the various, especially Westerly Yiddish, dialects whose testimony is crucial for the reconstruction of these items for Proto-Yiddish). Nonstandard forms (especially those in Alsatian and Swiss Yiddish) are usually given in the orthography of the primary source, at a considerable loss of consistency.

<sup>20</sup> While forms cognate with *koyletsh* occur in several German dialects, these are themselves borrowed from Slavic, either independently of Yiddish or perhaps indeed via Yiddish, but it does not seem possible that this word came into Yiddish via German. On the other hand, there are one or two other Slavicisms which Yiddish shares with some or all German dialects and which were undoubtedly borrowed into German early enough to have entered Yiddish as part of its German lexicon. Such words, e.g., *grenets* 'boundary' and probably *khreyn* 'horseradish', are ipso facto irrelevant to the issue at hand.

<sup>21</sup> Bin-Nun suggests a Hebrew origin: Hebrew *hataf* 'he grabbed' > \**khatf* > \**khapf* > *khap*-, but this does not seem nearly as likely as the Slavic origin, not the least because it is far from clear that there was a change of /pf/ to /p/ in Yiddish. And, even if there was, we would then still have to explain why this word shows up with /p/ even in those dialects which have the /pf/ phoneme.

<sup>22</sup> The Romance and Slavic elements I have in mind here are (some of the) ones which can safely be posited for Proto-Yiddish. On the other hand, there are, of course, numerous Slavicisms in Eastern Yiddish (or Easterly Yiddish generally) which are presumably considerably younger. As for the many Romance elements found only in "Western Yiddish" (or Westerly Yiddish specifically), the hypotheses presented in this paper would naturally tend to suggest that at least some of these are Proto-Yiddishisms which were lost in Eastern (or Easterly) Yiddish. However, more work is required on this point.

<sup>23</sup> What is at issue in particular is that we are dealing with words of uncertain etymology or else words made up of morphemes originally derived from different languages (and combined in Yiddish itself) or characterized by highly distinctive phonological, morphological, or semantic developments. These developments are characteristic of no language other than Yiddish, and if they are universal in Yiddish (or even merely attested at the extremes of the Yiddish dialectological spectrum), they can be considered as evidence of the unity of Yiddish and reconstructed for Proto-

'error, mistake';<sup>24</sup> *shekhtn* 'to slaughter', *mekn* 'to erase', and *rebe* 'rabbi';<sup>25</sup> *rebetsn* 'rabbi's wife';<sup>26</sup> *dalfn* 'poor man' and *šeygets* 'gentile lad';<sup>27</sup> *šikse* 'gentile girl';<sup>28</sup> *bal(e)bos* 'proprietor, owner; host; boss; master; landlord';<sup>29</sup> *hentshke* 'glove' and *hoyker* 'hump';<sup>30</sup> *fraynt* 'relative(s)', *kugl* 'kind of pudding', *gut-ort* '(Jewish) cemetery', *yortsayt* 'anniversary of death', and *shul* 'synagogue';<sup>31</sup> *yidishn* 'to circumcise';<sup>32</sup> *shabeyse-nakht(s)* 'Saturday evening' and

Yiddish. In the discussion that follows I try to provide some noteworthy examples of the various types of idiosyncratic Proto-Yiddish developments, but most types have many more instances than are listed here. In the notes, I try to point out what at least some of these Yiddish developments are, but only quite sketchily.

<sup>24</sup> While often listed as being of German and Hebrew origin, respectively, each of these words exhibits serious formal and semantic deviations from the putative prototypes in the source languages, and hence they are in effect indigenous to Yiddish.

<sup>25</sup> Words with Hebrew stems but with vowel changes due to (or at least suggestive of) German rules of umlaut, and other peculiarities.

<sup>26</sup> Derived from the Hebrew/German *rebe* by means of suffixes of debatable origin.

<sup>27</sup> Of Hebrew origin, but with meanings that are specifically Yiddish.

<sup>28</sup> Derived from *šeygets*, but not consistent with the rules of Hebrew/Aramaic word-formation, hence presumably proper to Yiddish.

<sup>29</sup> Derived from a Hebrew compound, but with an irregular phonological contraction of the second element. The feminine *bal(e)boste* is formed with the Aramaic suffix *-te*, also found in, e.g., *rošete* '[feminine] villain, wicked/malicious [woman]', *mekhutenešte* 'son-in-law's/daughter-in-law's mother; relative by marriage (fem.)', etc. If we allow the possibility that Yiddish was predated by a vernacular Judeo-Aramaic, as proposed by Katz (e.g., 1985), then it should presumably be possible to assume that such feminines made it into (Proto-)Yiddish from that language. This whole topic requires further investigation. However, Katz's hypothesis runs into such difficulties as the fact that, while Yiddish does have some Aramaic derivational morphology (such as notably the feminine *-te*), it seems to lack any Aramaic inflectional morphology, but does use the Hebrew plural *-im* and has some words with the (frozen) Hebrew dual *-ayim*. Given how languages usually seem to develop, this is more consistent with the hypothesis of Aramaic lexical borrowings into a Hebrew-based linguistic system than the other way around.

<sup>30</sup> Of obvious German origin, but phonologically irregular. Note that *hoyker*, together with *mezuze* 'mezuzah', are the sole examples given by Katz (1987) to demonstrate that some Yiddish words have idiosyncratic shapes as compared to their German and Hebrew sources, respectively. Specifically, their stressed vowels violate the general rules laid down by Bin-Nun and M. Weinreich for the correspondences of Yiddish stressed vowels to those in MHG and Hebrew-Aramaic. Although there is more to be said about these two cases (and about the additional examples mentioned in Katz 1986), Katz is quite right in general to emphasize the importance of Yiddish idiosyncracies as compared to its source languages. The only question is which of these idiosyncracies to reconstruct for Proto-Yiddish. Moreover, in any case, the case for the unity of Yiddish depends crucially on **how many** such idiosyncracies we can find. As even my (partial) list shows, they are in fact legion.

<sup>31</sup> Of straightforward German origin, but with meanings specific to Yiddish.

<sup>32</sup> One of many examples of a word made up of German-origin morphemes arranged

*umkheyn* 'disfavor';<sup>33</sup> *oysher* 'rich' and *mies* 'ugly';<sup>34</sup> *vayivrekht* 'flight';<sup>35</sup> and others too numerous to mention here;

(iv) the whole elaborate system whereby Yiddish derives verbs by combining Hebrew/Aramaic-origin stems (based on H/A inflected forms, usually the perfect or the active participle) with German-origin suffixes or auxiliaries; and such other morphological characteristics as the use of the German umlaut + *-er* plural with the Hebrew-origin noun *ponem/penimer* 'face(s)';<sup>36</sup> the verbal prefix *der-*; the abstract ending *-as*, as in *meshugas* 'craziness';<sup>37</sup> and so on;<sup>38</sup>

(v) a number of phonological developments such as the voicing of /s/ to /z/ in *muzn* 'must' and *lozn* 'to let'; the shortening of a long A-colored vowel (Proto-Yiddish A<sub>2</sub>) to /o/ (normally reflecting a Proto-Yiddish O<sub>1</sub>) in *lozn*, in some present-tense forms of 'to have', viz., 2sg. *host*, 3sg. *hot*,<sup>39</sup> as well as in some forms of Hebrew origin such as, e.g., *ho-(re)* 'the (evil)', *meshorsim* 'servants', and a few others;<sup>40</sup>

---

according to German morphological principles but in fact peculiar to Yiddish.

<sup>33</sup> Concatenations of Hebrew/Aramaic and German elements.

<sup>34</sup> Hebrew abstract nouns used as adjectives in Yiddish.

<sup>35</sup> A stem used in compound verbs such as Standard Yiddish *makhn vayivrekht* '(hum.) [to] run away, take to one's heels' [lit. 'make flight'], Swiss Yiddish *fiefrech houleche* [lit. 'walk flight']), etc. It derives from the Hebrew phrase *va-yivrax* 'and he fled' (*Genesis* 31:21).

<sup>36</sup> This may represent a remnant of a once more widespread pattern of using this German plural with Hebrew-origin nouns in *-im* (itself a plural ending in Hebrew) or in *-ayim* (the dual ending in Hebrew); compare Southwesternmost Yiddish *neelemmer* 'shoes', *einajemer* [i.e., *eynayemer*] 'eyes', *schinajemer* [i.e., *shinayemer*] 'teeth'.

<sup>37</sup> This apparently derives from what originated as Hebrew feminine sg. adjective forms of stems ending in a pharyngeal or laryngeal: Yiddish *meshugas* 'craziness' is directly related to Hebrew *meshuga* 'at' 'crazy (fem. sg.)'.

<sup>38</sup> Another potential Proto-Yiddish morphological feature worth mentioning may well be the use of the plural *-s* on a highly distinctive set of German noun stems, typically those ending in a vowel or in unstressed *-el*, *-er*, *-en*, *-em*, and *-ing*, something found only in a very restricted set of High German dialects (and in Low German and Dutch). This feature, if it really is Proto-Yiddish, would apparently be one of the few pointing to a Rhenish connection (Manaster Ramer and Wolf, in press). However, we have to be cautious, because the distribution of the *-s* plural in Western Yiddish is almost completely unknown.

<sup>39</sup> The /o/ in the 1sg. *hob* is a later, analogical development, apparently restricted to Eastern or Easterly Yiddish.

<sup>40</sup> Another possible Proto-Yiddish phonological development that bears touching on, although it is far from being worked out, is the lengthening of vowels in certain stressed final syllables (i.e., usually monosyllables). This is a particularly difficult problem, especially because there are all too few crucial examples. The generally accepted view appears to be that the lengthening we find must have occurred **before** the formation of Proto-Yiddish (i.e., in each of the source languages separately), but this has not been conclusively established. It does not seem impossible that the

--and so on.

In short, there is a whole host of lexical, morphological, and phonological peculiarities, all of which are found (certainly collectively but in most cases even individually) in no other language but Yiddish and at the same time are no less characteristic of Easterly (including specifically Eastern) than of Westerly (including specifically Southwesternmost, i.e., Swiss/Alsatian) Yiddish. Of particular interest perhaps are those features which are attested only in the westernmost and the easternmost Yiddish dialects, e.g., Swiss Yiddish *neerige* 'zu Tode schinden' ['to kill by flaying'] (Guggenheim-Grünberg 1976) ~ Ukrainian/Belorussian/eastern Polish Yiddish *nerik* 'maka, pega' ['wound, injury'] and its derivatives (U. Weinreich 1961:26),<sup>41</sup> *baa* 'Grossmutter' ['grandmother'] (Guggenheim-Grünberg 1976) ~ Piatra Neamt (Romania) Yiddish *bo* 'grandmother' (Herzog et al., to appear), etc.<sup>42</sup>

Even more telling than the lexical, morphological, and phonological facts may be the existence of a number of phrases or idioms which can be posited for Proto-Yiddish, e.g., *mishteyns gezogt* (an expression of pity or contempt depending on dialect but originally [and apparently still so in at least some "Western Yiddish" dialects] a formula used to avert bad luck (lit. 'may it be said to the stone'), see Herzog et al., to appear), *skotsl kumt* 'look who's here! welcome!' [usually addressed to women], *zayn der mer (mit)* 'to be the matter

---

lengthening in such words as *ov* 'fore-father, ancestor', *Ov* 'Ab, the 11<sup>th</sup> month in the Jewish calendar', *rov* 'rabbi', *kheyn* 'grace, charm, appeal', etc., was due to a phonological process, still to be analyzed in detail, which was active in Proto-Yiddish.

<sup>41</sup> It is particularly significant that this etymon (like the next one) involves some very specifically Yiddish developments. In particular, *nerik* (or rather *nerig-*) is an active stem in Yiddish, but it derives from a Hebrew passive verb form (probably the participle *neherag* '[one who was] killed'). Regarding the geographical distribution of *nerik* and related forms in Eastern Yiddish, U. Weinreich (1961) says that *nerik* occurs only in Ukraine, but that its derivatives are found also "bi-mkomot akherim" ['in other places']. Marvin Herzog (p.c.) informs me, based on unpublished data of the *Language and Culture Atlas of Ashkenazic Jewry*, that such forms are sparsely attested in Eastern Poland but widespread throughout both Ukraine and Belarus with the exception of a narrow border area between the latter two. Incidentally, this may be a good place to call attention to the striking fact that, while a Hebrew passive participle becomes an active stem in Yiddish in this case, the Hebrew active participle from the very same root becomes a passive in Yiddish (*hoyreg* 'killed person').

<sup>42</sup> This word, which reconstructs as Proto-Yiddish *\*bA<sub>3</sub>*, is of uncertain etymology. Even if (and this is by no means clear), it were related to Eastern (Easterly?) Yiddish *\*bA<sub>2</sub>A<sub>3</sub>be* (and hence to Slavic *baba*), as claimed by Wexler (1991:66, who in addition mistakenly lists *baa* as occurring in Alsatian Yiddish), it would be quite unclear just *how* it is related to the Slavic etymon. The loss of the second syllable would be a Yiddish innovation. Incidentally, since, as noted, Proto-Yiddish also clearly had *\*frO<sub>4</sub>le* 'grandmother', we must assume that *\*bA<sub>3</sub>* was a more intimate, perhaps nursery, term, one which we may want to gloss 'grandma'.

(with)', *esn teg* 'to eat as a guest at a certain house on given days of the week',<sup>43</sup> etc. The monogenesis of Yiddish thus appears to be amenable to as a clear and convincing a demonstration as that of any language or language family.

However, the work of establishing just which are the truly (originally) pan-Yiddish (or rather, since that is what really matters, Proto-Yiddish) features is far from done. For one thing, many of the phenomena often mentioned in the literature (e.g., Bin-Nun 1973:292-294 and passim, Birnbaum 1979:82-85 and passim, M. Weinreich 1973: passim) as illustrating the "fusion" of the different components of Yiddish and/or distinguishing Yiddish from any other language or (German) dialect are either clearly or at least likely restricted to some subset of Yiddish dialects, often (given the perhaps inevitable Eastern bias of much of Yiddish linguistics) Easternmost, Eastern, Easterly, or at least non-Southwesternmost Yiddish. Even such phenomena, discussed above, as the absence of /pf/ and /ç/, the devoicing of /b/ in nouns like *gopl* 'fork', *nopl* 'navel', etc., have all too often been mentioned as distinguishing Yiddish from German, even though they are not characteristic of all of Yiddish (or absent from all of German dialects). There are, of course, many more instances of this problem, too many to enumerate here, although some notable examples include the use of the Hebrew plural ending *-im* with the German-origin noun *poier* 'peasant'; the formation of certain feminine nouns in *-te*, such as *goyete* 'gentile, non-Jewish woman' (although the alternate form *goye* is of Proto-Yiddish date, as are some other feminines with the *-te* suffix, including *bal(e)boste*, *mekhuteneste*, and *roshete*); the semantic change in *vorem*, *vorn* 'because' (originally 'why'); and so on. Although often cited as setting Yiddish *tout court* apart from other languages and (German) dialects, these forms are in fact too restricted in their occurrence ("too eastern", so to speak, and hence too recent) to have belonged to Proto-Yiddish.

The complex issues that can arise in such cases can be illustrated with a closer analysis of one of M. Weinreich's (1973 [1980:609]) at first glance most striking examples of a supposed Yiddish peculiarity, namely, what appears to be the dual realization of Hebrew-origin *ḥa:ve:r* 'associate, companion, fellow' as both (a) *khaver* 'friend, companion, chum', with /a/ where /o/ would be expected according to the usual rules of correspondence between Hebrew-Aramaic and Yiddish vowels, and as (b) *khover* 'fellow, fellowship holder' [or rather a rank or title of a scholar], with the phonologically regular /o/ but a highly distinctive

<sup>43</sup> To this Standard (hence, Eastern) Yiddish form, compare Alsatian Yiddish *Tâg esse* (Weill 1920-1921, s.v. *Tag*). There is, of course, a problem concerning the precise reconstruction, since Eastern Yiddish has the plural form of the noun (lit. 'to eat days', whereas Alsatian Yiddish has the singular (lit. 'to eat day')). I would suspect that the Eastern Yiddish form, precisely because it is more "logical", may be a secondary refashioning of a prototype directly cognate with the Alsatian Yiddish form. Is it possible that the latter preserves a direct reflex of the (original) un-umlauted plural of this noun, which, due to the lautgesetzlich loss of the final schwa, would have ended up homophonous with the singular? The *Tâg* in *Tâg esse* would then be a reflex of *\*tA<sub>3</sub>g-e*.



semantic shift. However, while *khaver* presumably is indeed Proto-Yiddish, *khover* is likely to be a more recent and more local etymon (reflecting a relatively late borrowing from Hebrew into some Yiddish dialect or dialects which arose after the break-up of Proto-Yiddish, and thence into other Yiddish dialects). This is because, in Swiss Yiddish in particular it appears in a form which implies borrowing from another dialect rather than descent from a common Proto-Yiddish etymon. The crucial difference in history and chronology lies in the fact that, whereas Swiss Yiddish *chafer* has the expected /f/ reflex of Proto-Yiddish \*v, we find that *choower* has /v/, an irregularity best explained by borrowing from a more easterly Yiddish dialect (one in which \*v never devoiced). Thus, there was probably was no peculiarly Proto-Yiddish lexical split of a single Hebrew etymon into two different Yiddish ones such as M. Weinreich had in mind, since the etymon of *khover* was not yet found in Yiddish at the time.

To be sure, something peculiar must have happened in Proto-Yiddish to explain the /a/ of *khaver*, but here the issues become even more complex. M. Weinreich considered the /a/ of *khaver* to be just one of a largish number of instances of /a/ (Proto-Yiddish A<sub>1</sub>) in Hebrew-Aramaic words where /o/ (Proto-Yiddish A<sub>2</sub>) is expected. He sought to explain all these examples as representing a special layer of Hebrew-Aramaicisms which entered Yiddish earlier than the rest of its Semitic vocabulary. However, almost all of these "unexpected" /a/'s appear in closed syllables, an environment in which A<sub>1</sub>, and not A<sub>2</sub>, is in fact regular (Bin-Nun 1973:271, Katz 1986 and passim). This, together with other facts which I will not go into here (see, e.g., Manaster Ramer, to appear b), means that the whole theory of the two different layers of Hebrew-Aramaic vocabulary in Proto-Yiddish is probably unnecessary.<sup>44</sup> It thus becomes imperative to find another explanation for *khaver*, which now stands almost alone. One possibility is that the /a/ in *khaver* would not be irregular after all if this word did not, as is usually assumed, derive from the same Hebrew etymon as *khover*, but rather, as posited by Weill (1920-1921, s.v. *Hawar*), from the synonymous Aramaic *ḥavar* instead. The Proto-Yiddish peculiarity in that case would be, not the vowel of *khaver*, but rather the (putative) suppletive relationship which would exist between *khaver* (if this is really Aramaic) and its plural, *khaveyrim*, given that the latter clearly comes from the Hebrew *āve:rim* (where the /a/ is regular). There would thus be a Proto-Yiddish peculiarity, though not the one that M. Weinreich was assuming.

Nor is this the end of the story, for it is also possible that the /a/ vowel of *khaver* is not a reflection of a putative Aramaic origin at all, but (as proposed by Bin-Nun 1973) that it arose via analogical leveling with the plural, where as noted the /a/ is regular. On this scenario, it would be precisely this analogical development which would represent the Proto-Yiddish peculiarity in this word.

<sup>44</sup> I should point out that the theory of "older" and "more recent" layers of Semitic in Yiddish was held not just by M. Weinreich. Bin-Nun used the same theory to explain some other problematic examples, although, as noted, not the /a/'s in closed syllables--or in *khaver*.

Moreover, since there are cases in Yiddish where Hebrew-Aramaic *ă* yields /o/ rather than /a/ (and no one has yet figured out the rules for this), it is even possible that Aramaic *ḥavar* could be the source of *khover* rather than of *khaver*--and it appear that it is the Aramaic, rather than the Hebrew, semantics that explains the meaning of *khover* (Robert Hoberman, p.c.). If so, then we would have a reversal of what was proposed by Weill: *khaver* would be of Hebrew origin (with /a/ of analogical origin), while *khover* would be Aramaic. Clearly, either we need more work or, perhaps, we may have to conclude that the issues are undecidable (and focus on other, easier problems).

Even greater complexities are involved in evaluating Katz's (1983, 1987) claim that the unity of all (or, if I read the later work correctly, of almost all) of Yiddish follows from the fact that different Yiddish dialects exhibit the same pattern of fusion of the different components of the Yiddish lexicon (German, Romance, and Hebrew-Aramaic, although I would have to add Slavic) as well as a system of interdialectal vowel correspondences consistent across these components (the very correspondences which, of course, constituted the basis of the "protovowel" system proposed by M. Weinreich). As should be apparent even from the few issues discussed in this paper, the real situation is far from this simple. First, as far as the fusion of the four components of the Yiddish lexicon is concerned, the lexical isoglosses between Eastern (or Easterly) and Western (or Westerly) Yiddish which were discussed above indicate that there are, after all, major differences among Yiddish dialects with regard to the choice of German-, Romance-, Hebrew-Aramaic-, or Slavic-origin words (differences which can only be explained by assuming that the processes of lexical selection went on long after the Proto-Yiddish period). Second, as for the correspondences among the vowel systems of the Yiddish dialects, there are both factual and logical difficulties with Katz's argument. On the logical side, Katz's argument appears to demand that (M. Weinreich's) proto-vowel system only work for those components of the Yiddish lexicon which had "fused" at the very inception of Proto-Yiddish. Yet, on the one hand, Katz himself (like M. Weinreich) does not accept the existence of a Slavic component in the Proto-Yiddish lexicon, but, on the other hand, M. Weinreich's proto-vowel system extends, as the latter has shown in detail, to the Slavic component of (Eastern) Yiddish. Hence, there would seem to be an internal inconsistency in Katz's argument. On the factual side, there are, especially in the Westerly Yiddish dialects, several deviations from the rules posited by M. Weinreich besides the problems, discussed above, with the reflexes of  $E_4$ ,  $O_4$ , and  $\ddot{O}_4$  (with many more probably waiting to be discovered). As a result, it is **not** the case that all the vowels of all the Yiddish dialects are as of now accounted for, and what is crucial, it is not yet clear whether the vowels of Hebrew-Aramaic-origin as opposed to German-origin words really do behave the same across the dialects. It is striking in this context that Zuckerman (1969) more than once specifies **different** Alsatian Yiddish reflexes for certain Yiddish proto-vowels depending on whether they occur in words of German or Semitic origin. Although I do not think he is right, no one really knows at present, and certainly the problems need to be solved before we can

venture the kind of argument Katz has advanced. In both lexicon and phonology, then, a simple juxtaposition of the different synchronic dialectal systems does **not** argue all that strongly for the unity of Yiddish. But that is precisely the point: what we want is more than mere juxtapositions of the synchronic dialects.

Instead, we must aim at a genuine reconstruction of Proto-Yiddish (lexicon and phonology as well as phraseology and grammar). The (partial) list of lexical, morphological, phonological, and phraseological items given above to illustrate the kind of argument that can be given for the original unity of Yiddish also serves to illustrate the kinds of things that we can, and must, try to reconstruct for Proto-Yiddish. We can only hope to demonstrate that all of Yiddish comes from a single Proto-Yiddish if we can show that this (hypothetical) linguistic system was characterized by a historically unique pattern of phonological, lexical, and grammatical developments distinguishing Yiddish from its source languages. The resulting picture of the origins of Yiddish may perhaps end up being quite similar to current views in broad outline, but the many specific changes which are certain to be required as we gradually work out a historically realistic reconstruction<sup>45</sup> will mean the difference between merely having a general idea that the different Yiddish dialects **can** be more or less systematically related to each other--and being able to demonstrate that these dialects in fact **did** emerge from a single, historically unique, proto-language.

9. Once arguments such as those alluded to above are fleshed out (which is a biggish, but both conceptually and factually quite straightforward, undertaking), the original unity of Yiddish can be considered established. However, both the reconstruction of many other aspects of Proto-Yiddish and the subsequent history (and the family tree) of Yiddish dialects will still require much additional work. Nor can we even begin to guess where Proto-Yiddish was spoken, or (beyond the little that was said here) much about the locations of the earliest Yiddish dialect divisions. The issues here are complex, although by no means hopeless.

A good example, where the answer is still far from clear but progress is clearly being made, is that, contrary to conventional wisdom (e.g., M. Weinreich 1973, Katz 1983:1018), but in agreement (to this small but nontrivial extent) with Wexler (1991:65-72), I have to hypothesize that Proto-Yiddish must have had a few Slavisms (*nebekh*, *koyletsh*, *khotsh(e)*, *khapn*, and perhaps no others). What is crucial here is, not only that there are Slavisms in all Yiddish dialects, but that the same small set is found throughout all the Westerly varieties,<sup>46</sup> and that this set is entirely different from the collection of Slavic

<sup>45</sup> The (at a superficial first glance, quite minor) revisions of M. Weinreich's vowel system in Katz (1983) mark a significant conceptual shift away from diaphonemics and towards true reconstruction.

<sup>46</sup> Further east, there are of course, many more Slavisms, but these are later and hence irrelevant. Note in particular that the vast majority of the Slavisms which Wexler (1991:65-72) posits for the earliest stages of Yiddish are demonstrably more recent, being restricted to (subdialects of) Eastern Yiddish or at best Easterly Yiddish. One example will suffice here: although he claims that *kachke* 'duck' is "attested

loanwords to be found in any attested variety of German. Thus, this set of Slavisms almost certainly must be a characteristic of (Proto-)Yiddish itself, indicating that some kind of influence from the east must be assumed for Proto-Yiddish. (contrary to the classic Rhenish theory of authors such as M. Weinreich). What this means more specifically for the genesis of Proto-Yiddish, it is still too early to tell. For one thing, there are so few of these Slavisms that an origin in eastern Germany (as claimed by the Danubian theory of Katz, Faber, and King) or indeed in Slavic lands (as posited by Wexler) is not strongly indicated either. In particular, this handful of Slavisms does not suffice to prove a relationship with a hypothetical Judeo-Sorbian, as claimed by Wexler. In fact, it is not even clear that these words must come from any kind of Judeo-Slavic at all, as opposed to from Old Sorbian (or perhaps Old Czech). Unlike in the case of the Romance vocabulary in Yiddish, which cannot be derived directly from any of the (Christian) Romance languages and clearly points to a specifically Jewish source (a Judeo-Romance or Judeo-Latin of some kind), the Slavisms in Yiddish, as far as I can see, can be derived straightforwardly from Old Czech or Old Sorbian, and there is as yet no indication of specifically Jewish developments requiring us to posit a Judeo-Slavic source in these cases. A fortiori, I see no basis in fact for Wexler's claims about the origin of the whole of Yiddish as a Judeo-Slavic language which was then relexified with Germanic vocabulary.<sup>47</sup>

Another example, where the answer seems rather clearer despite the complexities of the data, involves the few but striking features shared by some Yiddish dialects with Austro-Bavarian dialects of German. Contrary to the King-Faber-Katz theory of a Danubian (i.e., Austro-Bavarian) origin of Yiddish, Manaster Ramer and Wolf (in press) argue instead for an Austro-Bavarian influence on some early form of Central Yiddish (the westernmost division of Eastern Yiddish) along with some immediately adjacent easterly Western Yiddish dialects (e.g., those of Bohemia). This means that the Austro-Bavarian connection (to the extent that it is real at all, for many of the phenomena cited by King et al. are provably independent developments in Eastern Yiddish dialects and in Austro-Bavarian German or else are features common to much of High German) occurred long after the break-up, not only of Yiddish, but even of Easterly Yiddish and of Eastern Yiddish, into individual dialects.

In spite of all the work that has been done, we still have more gaps than filled-in areas in our knowledge of Proto-Yiddish and the early history of the dialects descended from it. The results reported here, coming roughly a century after the birth of comparative Yiddish linguistics, are only scratching the surface

---

throughout Yiddish", in reality it is not even pan-Eastern Yiddish and is clearly a late innovation of the (majority of) Eastern Yiddish dialects, replacing the Germanic etymon *entl*.

<sup>47</sup> Not to mention the fact that it would also have to be assumed that this language was "regrammaticalized" (as well as "relexified"), inasmuch as there is no trace of Slavic grammatical influence in any reasonable reconstruction of Proto-Yiddish. It is perhaps not merely a quibble to ask what sense it makes to say, of a language whose lexicon and grammar are both of non-Slavic origin, that it is a Slavic language.

of what promises to be a major research topic in the next century. As in every area of comparative linguistics, whether on the large scale as in the case of such hypotheses as Nostratic, Altaic, Na-Dene, etc., or on the smaller scale, as in the case before us, more work is called for.

### References

- Beem, H.  
 1959 *Jerösche: Jiddische Spreekwoorden en Zegswijzen uit het Nederlandse Taalgebied*. Assen: Van Gorcum.
- Beranek, Franz J.  
 1961 "Die fränkische Landschaft des Jiddischen". *Jahrbuch für fränkische Landesforschung* 21: 267-303.  
 1965 *Westjiddischer Sprachatlas*. Marburg: N. G. Elwert.
- Bin-Nun, Jechiel  
 1973 *Jiddisch und die deutschen Mundarten unter besonderer Berücksichtigung des ostgalizischen Jiddisch*. Tübingen: Max Niemeyer.
- Birnbaum, Solomon A.  
 1954 "Two Problems of Yiddish Linguistics". *Field of Yiddish* 1: 63-72.  
 1979 *Yiddish: A Survey and a Grammar*. Toronto and Buffalo: University of Toronto Press.
- Garvin, Paul L.  
 1965 "The Dialect Geography of Hungarian Yiddish". *Field of Yiddish* 2: 92-115.
- Guggenheim-Grünberg, Florence  
 1965 "Place Names in Swiss Yiddish". *Field of Yiddish* 2: 147-157.  
 1973 *Jiddisch auf alemannischem Sprachgebiet*. Zurich: Juris Druck + Verlag. [Beiträge zur Geschichte und Volkskunde der Juden in der Schweiz, 10.]  
 1976 *Wörterbuch zu Surbtaler Jiddisch*. Zurich: Juris. [Beiträge zur Geschichte und Volkskunde der Juden in der Schweiz, 11.]
- Herzog, Marvin I.  
 1969 "Yiddish in the Ukraine: Isoglosses and Historical Inferences". *Field of Yiddish* 3: 58-81.  
 To appear *Language and Culture Atlas of Ashkenazic Jewry*, III.

- Hutterer, C. J.  
 1965 "The Phonology of Budapest Yiddish". *Field of Yiddish* 2:116-146.
- Katz, Dovid.  
 1983 "Zur Dialektologie des Jiddischen". In: Werner Besch et al. (eds.), *Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialekt-forschung*, 1.2: 1018-1041.  
 1985 "Hebrew, Aramaic and the Rise of Yiddish". In: Joshua A. Fishman (ed.), *Readings in the Sociology of the Jewish Languages*, 85-103. Leiden: Brill.  
 1986 "Ha-yesod ha-shemi ba-yidish: yerusha mi-ymey kedem". *Ha-sifrut* 3/4 (35/36):228-251 [English abstract, p. viii].  
 1987 "The Proto Dialectology of Yiddish". In: *Origins of the Yiddish Language*, ed. Dovid Katz [= *Language and Communication* 7, Supplement], 73-81. Oxford: Pergamon Press.
- King, Robert D.  
 1987 "Proto Yiddish Morphology." In: *Origins of the Yiddish Language*, ed. Dovid Katz [= *Language and Communication* 7, Supplement], 73-81. Oxford: Pergamon Press.  
 1992 "Migration and Linguistics as Illustrated by Yiddish". In: Edgar C. Polomé and Werner Winter (eds.), *Reconstructing Languages and Cultures*, 419-439. Berlin, New York: Mouton de Gruyter
- King, Robert D. and Alice Faber  
 1984 "Yiddish and the Settlement History of Ashkenazic Jewry". *The Mankind Quarterly* 24: 393-425.
- Landau, Alfred  
 1895 "Das Deminitivum der galizisch-jüdischen Mundart". *Deutsche Mundarten* 1: 46-58.
- Landau, Alfred, and B. Wachstein  
 1911 *Jüdische Privatbriefe aus dem Jahre 1619*. Wien, Leipzig.
- Lowenstein, Steven  
 1969 "Results of Atlas investigations among Jews of Germany". *Field of Yiddish* 3: 16-35.  
 ms. "The Shifting Boundary between Eastern and Western Jewry".
- Manaster Ramer, Alexis  
 To appear a "gayes".  
 To appear b "How selective was Proto-Yiddish?". *American Journal of Germanic Linguistics and Literatures*.

- Manaster Ramer, Alexis, and Meyer Wolf  
 To appear "Yiddish Origins and the Austro-Bavarian Problem". *Folia Linguistica Historica*.
- Marchand, James W.  
 1987 "Proto Yiddish and the Glosses: Can We Reconstruct Proto-Yiddish." In: *Origins of the Yiddish Language*, ed. by Dovid Katz [= *Language and Communication*, 7, Supplement], 83-94. Oxford: Pergamon Press.
- Prilutski, Noyakh  
 1920 *Tsum Yidishn Vokalizm*, I. [= *Yidishe Dyalektologishe Forshungen* 4]. Warsaw: Nayer Ferlag.
- Shpirn, Tsvi  
 1926 "Di Yidishe shprakh in Ungarn". *Filologishe Shriftn* 1:195-200.
- Timm, Erica  
 1987 *Graphische und phonische Struktur des Westjiddischen*. Tübingen: Niemeyer.
- Weill, Emmanuel  
 1920-1921 "Le yidisch alsacien-lorrain". *Revue des études juives* 70:180-194, 71:66-88, 165-189, 72:65-88.
- Weinberg, Werner  
 1969 *Die Reste des Jüdischdeutschen*. 2d ed. Stuttgart: Kohlhammer.
- Weinreich, Max  
 1940 "Yidish". *Algemeine Entsiklopedye*, vol. Yidn Beys, 91-100. Paris: Dubnov-Fond.  
 1958a "Bney-hes un bney-khes in Ashkenaz: Di problem--un vos zi lozt undz hern." In: *Shmuel Niger-bukh* (ed. Shlomo Bickel and Leibush Lehrer), 101-123. New York: Yivo.  
 1958b "Roshe-prokim vegn mayrevdikn yidish". In: *Yuda A. Yofe-bukh*, 158-194. New York: Yivo.  
 1973 *Geshikhte fun der yidisher shprakh*. New York: Yivo.  
 1980 *History of the Yiddish Language*, tr. by Shlomo Noble, with the assistance of Joshua A. Fishman Chicago and London: University of Chicago Press.

## Weinreich, Uriel

- 1961 Ha-ivrit ha-ashkenazit ve-ha-ivrit she-be-yidish: bekhinatan ha-geografit (hemshekh). *Leshonenu* 25: 57-80
- 1964 "Western Traits in Transcarpathian Yiddish". In: *For Max Weinreich on his Seventieth Birthday*, 245-264. The Hague: Mouton.
- 1968 *Modern English-Yiddish Yiddish-English Dictionary*. New York: Yivo and McGraw-Hill.

## Weiss, C. Th.

- 1896 "Das Elsässer Judendeutsch." *Jahrbuch für Geschichte, Sprache und Literatur Elsass-Lothringens* 12:121-182.

## Wexler, Paul

- 1991 "Yiddish, the Fifteenth Slavic Language. A Study of Partial Language Shift from Judeo-Sorbian to German." *International Journal of the Sociology of Language* 91: 9-150.

## Zuckerman, Richard

- 1969 "Alsace: an Outpost of Western Yiddish". *Field of Yiddish* 3:36-57.



## Etymological Problems with Words for 'Blood' in Nostratic and Beyond

Karl Heinrich Menges  
University of Vienna

The common Tungusic word for 'blood' is *se-kse*, with a number of variants and derivatives. The simplex *\*se:-* can well be considered as the Tungus root and occurs, as far as hitherto known, as a verbal base only: Negidal *se:-* 'to bleed' (intrans.). Of the numerous variants listed in Cincius (1975-77, II:138, the following selection shall suffice: Evenki *se:-kse* (Stony Tunguska, Baunty, Ne:pa, Ušami, Saxalin, Tokko, Urmi, Učuri, Čumykan), *se:γse* (A), *se:-hse* (Stony Tunguska, Tokma), *se:-he* (Aldan, Upper-Lena, Kačug, Tokma, Tommot, Xingan), *ho:-kse* (Agata), *he:-kse* (Dudinka, Viljuj, Jerbogačen, Ilimpija), *še:-kše* (Sym, North-Bajkal), *še:-wše* (St. Tunguska, North-Bajkal), *še:-kše* (Sym), *še:-he* (St. Tung., Tokma), etc. 'blood', *se:-γdi* (St. Tung., Ne:pa, Urmi), *še:-di* (Sym) 'bloody, bloodstained', *se:-kse-gde*, *se:-hsegde* (St. Tung.), *se-hegde* (Kačug) 'bloody', *se:-kse-re:-w-*, *se:here:w-*, *se:re:w-* (many central dialects) 'to become bloodstained', *he:-re:-ptin* (Ilimpija) 'sacrificial blood (usually of reindeer)', *ho:-kše-ti-* (Agata) 'to drink blood (of a killed animal)'; Solon: *se:-kče*, *se:-tče* 'blood'; Lamut: *he:s* (< *se:-kse*) 'bloodcrust'; *he-du-*, *hedu-l-* 'to get subcutaneous bloodstaining from contusion, etc.'; Armañ: *se-s-mi-* 'red(dish) color, tint' < *se:-kse-mi:-*; Oroči: *se:-kse* 'blood'; Udihe: *sakeä* 'id.'; Olča: *se:-kse* 'id.'; Oroki: *še:-kše* 'id.'; Na:naj: *se:-kse* 'id.', *se:-l*, *se:-l-ži-ni* 'convulsions, fits of pain'; Manžu: *se-ŋgi* 'blood'; Žürčen: *se-ŋi* 'id.'.

Besides these rich formations in Evenki, Širokogorov (1944, II:38) offers *se:kse* (Birarčen, Nerča), *šakša* (Xingan), with *š* not explained in his work, but apparently meaning a sound resulting from the transition from original Tungusic *s-* > *h-*, as observed by Vasilevič (1948: 161) in the Ilimpija dialect. Further he lists *soksä* and *säksä* (Manegir, Urulga [Castrén]), *sjöksjä* for Negidal (after Schrenk) and Na:naj (Goldi), *söksö* (also from Schrenk), and *sö-gi* from Žürčen (as above); those he compares with Nivx (Giljak) *čeo*x and Mongolian *čisun*; he continues to list, but separately, *sü:ksa* for Birarčen (with *š*, apparently = *ś*), with a vague, not well legible comparison with Russian *sok* 'sap, juice' and Lamut *xuŋyel* (dial. of Lamunxan) and *xuŋol* (dial. of

Tumunxan), 'blood'; finally he gives for 'sacrificial blood' *aiga*, *ajga sa:kša* for Birarčēn, as a religious term and 'clotted blood', *nu:ži* (Manegir) with reference to Buřat *nuži* 'id.'. Of these latter words the two Lamut forms are well-known (cf. *infra*), while *aiga* and *ajga sa:kša* seem to be a derivative of *aj* 'help' (Cincius, 1975-77, I:17) like Ew. *ajgan* 'medication', *aj-i:-či-mñi* 'healer, physician' and Lamut *aj-n-u-*, *aj-ŋ-u-ru-* 'to repair, retribute, straighten out, revive, heal' (Cincius, 1975-77, I:17), all from *aj* 'help', *aj-*, *aj-y-* 'to help' (ib.), together with many other derivatives.

While those may serve as examples of the rich derivational forms, from the other Tungus languages only the basic lexemes have been replaced here. Only two of all the Tungus languages have no derivative of *se:-* for 'blood', like the most common *se:-kse* (and variants), but a different etymon, *huŋel*, *xuŋge:l* in Lamut and *suŋgo:l* in Armañ (cf. Cincius & Rišes, 1952:238; Cincius, 1975-77, II:350) whose etymology is unknown so far; as a loanword from Lamut, *huŋel* has been found in Evenki in the dialect of Ngokonno only (ib.). The regular Lamut form of *se:-kse* is *he:s*, preserved only with the special meaning of 'blood-crust, clotted blood'. Elsewhere in Tungus, the ancient root *se:-* is ubiquitous. It might be compared, in view of animistic and shamanistic belief with Mongolian *süne-sün* 'soul', which also occurs as a loanword in Evenki, *sunesun* ('id.'), so far established by Castrén only for the Ev. dialect of Nerča, but the etymological comparison of the two lexemes presupposes a Tungus equivalent of the etymon underlying Mong. *sünesün*, which has not yet been found. The important role that blood plays with all animistic and shamanistic beliefs and its various rites should have lead, especially in Altaic, to the creation of a considerable number of taboo-expressions, even within one and the same subgroup such as Tungus, but there is no evidence of it. The situation in Indo-European is just the opposite: there, even one and the same single language of a subgroup, such as e.g. Latin and Greek, may have more than one lexeme at their disposal, as e.g. Latin *sanguis* and *cruor*, or Greek *αἷμα*, *ζαῖρα*, *ἰχὼρ*.

The Altaic family with its five branches does not possess a common etymon for 'blood'. Each has its own term: Turkic *qan* (< *qa:n*), Mong. *čisun* (< *\*ti-sun* < *\*ty-sun*), Kor. *p'i*, Jap. *či* (< *\*ty/ti*). This may well be due to an earlier variety of lexemes, often caused by taboo, of which only a few, one in each group, have survived as the basic term. This development may likewise have taken place in Nostratic.

Here, the original etymon underlying later Altaic Tungus *se:* is extant, with great probability, in the IE root *\*sei-/soi-* 'to drip, drop, run (liquids); humid, wet' (Pokorny, 1959:889), cf. the semantic parallel in Dravidian, Tamil *čo:r* 'id.', and its large family (Burrow and Emeneau, 1961: no. 2883). To the former, Pokorny connects M-Irish *silid* 'drips, drops, runs', Cymr. *hufen* 'cream', OHG, NHG *seim* '(pure) honey' (< *\*soimeno-*) that have their exact

correspondence in Tungus *se*: < \**soi-/sai-*, the Tungus *e*: resulting from Proto-Altaic \**ai/oi*. Pokorny lists neither Lat. *sanguis*, nor Gk. αἷμα, which are not even mentioned in the entire Etymological Dictionary. An IE derivative from \**sei-/soi-* might possibly underlie αἷμα from an earlier \**sai-m-nt-*, a periphrastic formation that originated under taboo. In search of an etymology of αἷμα it was compared with Old Icelandic *hunangs-seimr* 'Honigseim', OHG, NHG *seim* 'id.', Later NHG *Seim* < IE \**soi-meno-* "cream" (cf. Boisacq, 1950:24, according to whom this comparison was rejected by Kluge, erroneously, as I think; IE \**soimeno-* is morphologically to be analysed as \**soi-m-en-o-*, \**soi-m-nt-* [sai-: αἷ-]). A number of other etymologies was offered, l.c. and by Walde-Hofmann (1938, 1954, II:474f. sub *sanguis*). The latter authors remark: "Herkunft unklar", and they continue with the correct statement that "Die Wörter für 'Blut' differieren von einer Sprache zur anderen (vgl. z.B. Gr. αἷμα, Goth. *bloþ*, Gr. ἱχὺρ 'Götterblut, Blut')". The traces of the development have been obfuscated by taboo which generally does not result from a mere substitution of one etymon by another, but produces willful, arbitrary and artificial alterations made in pre-literary times or in non-literary circumstances. For this, a look into shamanist texts, even quite recently recorded ones, with their ever recurring tabooistic obumbrations would prove both instructive and sufficient. Thus IE Gk. αἷμα, Lat. *sanguis* and Altaic Tungus *se*: originally are taboo expressions, arcane circumscriptions with the basic semantics of 'dripping, running, streaming, trickling', as profusely represented by the above Dravidian family of (Tamil, etc.) čo:r (Burrow and Emeneau, 1986: no. 2883), that renders illusory all attempts to reach irreproachable etymologies.

The number of the basic etyma for 'blood' in the various Nostratic groups differs from one to the other considerably, usually in accordance with the number of subgroups (or families), so that for Nostratic at least six may be posited, for IE more than a dozen, while Dravidian has, according to those listed so far in Burrow and Emeneau (1986), ten, although not consisting of an equal number of subgroups. Uralic seems to have one for Finno-Ugric, but at least two for Samoyedic that exhibit some relationship with Turkic and with Nivx (Gilyak, cf. infra). Inasmuch as they are recognizable so far, the Dravidian lexemes seem to be descriptive, in some instances metonymic like those in IE, a fact without doubt due to taboo that has prevented the use and development of basic concepts, and this in the course of long historical duration. The same situation is to be assumed for Uralic and Altaic, but in view of the much smaller number of etyma, the linguistic evidence still is not clear enough. For the time being, material from Kartvelian and the vast groups of AfroAsiatc have not yet been examined.

But Proto-Altaic \**se*:, Nostr. \**sei-/soi-* has an intriguing parallel far outside of Nostratic, in the Tibeto-Burmese subgroup of Sino-Tibetan: Burmese *swè*, Hor *sje*, Tangut *šie*, all meaning 'blood' (cf. Shafer, 1960:164, and

Nevskij, 1960, I:339. Nevskij compares this etymon with A-hi Lolo *sō* 'id.'). This latter Tibeto-Burmese and Tangut etymon for 'blood' can hardly be separated from its Chinese equivalent, *süä* 血 < M.-Chin. *xiwet* < A.-Chin. *\*xiwet* in Karlgren's (1940: No. 410, a-c) reconstruction, in both cases apparently disyllabic, while Shafer assumes (l.c.) A.-Ch. *\*x'weð*. The final dental of A.-Ch. and M.-Ch. survives in Kanton *hüt*, Hakka *het*, and its forms in Annamite, *hüet*, Japanese *keccu*, *keči*, and Korean *hjöŭl* (Wade-Giles, no. 4847). The oldest Chinese written form of *\*xiwet* exhibits the intimate connection of the idea of 'blood' with that of a sanguinary sacrifice, as seen in no. 410 b and c, quoted and explained by Karlgren, l.c., as that on a Jin oracle bone, b and c, found in an inscription of C'zou II: "The graph is a drawing of a sacrificial vessel with content". This immediately reminds one of the semantics of the ἄπαξ εἰρημένον in Odyssey, III, 444, ἀμύλον 'vase où l'on recueillait le sang de la victime' (Boisacq, 1950:54) which Wilhelm Schulze connected with Lat. *sanguis*, derived by him from a postulated IE *\*sangʷen* via *\*σαμβν-* which was doubted by Meillet, Osthoff and Boisacq, and rejected by Walde-Hofmann (1938, 1954, II:474 f.). It was just one of the many etymological constructions rendered futile by taboo. Maybe, Schulze was nevertheless on the right path? The intimate relationship of blood, soul, and life in many cultures equally existed in European Classical Antiquity, as seen e.g. in Hesychios' gloss: ῥαρ αἵμα, ψυχῇ (cf. Boisacq 1954:209 sub ἔαρ, <sup>1</sup>).

For common Tungusic *se-*, Proto-Altaic and Proto-IE *\*se-*, *\*sei-*, *\*sai-*, *\*soi-* no equivalent cognate etymon has been found so far in East-Nostratic; the search still has to be extended to Kartvelian and Afroasiatic. In view of their relatively far-reaching occurrence, a few lexemes for 'blood' of Nostratic and extra-Nostratic origin shall be mentioned here. As to Uralic, of the Samoyed lexemes for 'blood', as listed in Castrén and quoted supra, the most repanded one seems to be Neneč (Jurak) *xeam*, *he:m*, Nganasan (Tavgy) *qam*, and Sölkup (Ostjak) Narym *qap*, Ket' *qam*, Čulym, Upper Ob *käm*, Jeloguj, Bajxa, Taz, Karaš *këm*, and in Kamaš *k'em* (which is also quoted in Donner (1944:28, 29): *k'em*, *k'ëm*; not in Joki's "Lehnwörter des Sajansamojedischen" as Joki does not consider it a loanword but a native form). None of the Samoyed words has final *-n*, but only a final labial, *-m*, in Sölkup of Narym *-p*. Proto-Tk., Ka:šyari: and Türkmen have *qa:n*, Jakut *xa:n*, common-Turkic. *qan*, Tavaš *jun* 'blood'. The Samoyed etymon has a surprisingly close counterpart in Ainu, *kem* 'blood' (cf. Hattori, 1964:19 f., nos. 162, 164). These two etyma, Samoyed *qam/käm* etc. and Ainu *kem* transcend the boundaries of Nostratic and demand thorough further investigation as they might well be considered as traces of a more distant ancient genetic relationship.

In the "narrower" field of Altaic, traces of Tk. *qa:n* are found in Tungus, preponderantly in the North: Evenki Učur, Urmi, Saxalin *haŋa-* 'to bleed (intr.)', Ilimpija, Sym, Učur, *haŋi-*, Ilimpija, Norboko:, Urmi, Saxalin *ha:ŋe* 'to drink

the blood of a killed game', Lamut (most dialects) *haŋ-* 'to bleed (intr.)', Olja *ha:ŋeʒaq* (and many derivatives in the dialects) 'bleeding', Oxotsk *ha:ŋtar-* (numerous derivatives in the dialects) 'to stop bleeding, to cease (of bleeding)', Negidal *xana-* 'to bleed (trans.), take blood', Udi *xana-ʒa-* 'id.' (Cincius, 1975-77, I:372; Vasilevič, 1958:469). Those will be dealt with later.

From Jenisejic Castrén (1858:187, 234) noted Ket *sul*, *su:l*, *śul*, *śu:l*, and Kot *šur* 'blood', which cannot be separated from Turkic Ujgur *söl*, Ka:šyari: *sö:l*, Qazan-Tatar *sül*, etc. 'juice of fruit, meat; soup, bouillon' also 'humidity, humid excretion, pus'. Räsänen (1969:430) has no etymology, but a reference to Čayatai *sülägäj*, Mongolian *silükej* 'spittle', Tungusic Evenki, etc. *sile-kse* 'dew', Finnish Suomi *sykki* 'spittle', of which the Turkic word seems to be a loan from Mongolian derivatives from the same stem or root. Mongolian has, e.g. in the Jüan-C'ao *Bi-Sy šilän*, Lit. Mong. *šile*, *šüle*, Xalxa *šölö*, in Tungus Manžu *šula* 'fruit-juice, opaque liquid', *sila* 'soup'; this apparently a loan from Mongol, and Korean has *sul* 'wine, brandy'. The Altaic etymon furthermore points to Indo-European Lithuanian *sulà* 'sap of birch (rarely other trees)', Russian *Sulá*, the name of five rivers quoted in Vasmer (1950-59, III:43), compared there with Greek ὕλη 'Urstoff, primordial matter; morass, mud'. This is connected by H. Güntert with Old Indic *śura:*, A:vestan *hura:* 'brandy', by others with reference to 'birch sap, fruit juice', with Old-Indic *su-no:ti* '(he) presses (out)', which is less probable but demanding further investigation, just as Romance French *souiller*, Provençal *solhar*, Catalan *sulhar* 'to soil, dirty' (cf. Meyer-Lübke 1968, no. 8418) and Germanic NHG *Suhle*, *suhlen* 'to roll, wallow in the mire' (cf. Kluge and Mitzka, 1975). If the Romance lexemes result, as Meyer-Lübke thinks, from Latin *suculare*, they do not belong here. If the above Jenisejic etymon is not of genuine Jenisejic origin, it might very well be due to an ancient borrowing from Turkic *söl* in order to satisfy the tabooistic needs of Jenisejic peoples who at the times prior to the Russian conquest must have consisted of a greater number of speakers and have been living in closer vicinity with Turkic peoples. Likewise, the probability of a far distant relationship of Altaic *söl*, Indo-European Baltic and Slavic *sula*, Old-Indic *śura:*, A:vestan *hura:* with the Jenisejic words and Chinese *\*xwe:t*, *\*xwe:ð*, *śüä* might be taken into consideration here.

In a recent discussion of the above Tibeto-Burmese etymon and Chinese *śüä* < M.-Ch. *\*xiwet* < A.-Ch. *\*xiwet* (< *\*xweð*, after Schaefer), Roy A. Miller came to the conclusion that the Chinese etymon had original, proto-Chinese *\*x-*, and therefore is to be considered as a different etymon, not belonging to the above Tibeto-Burmese, nor to the Tungus series, with their initial *s-*. However, I now think it is noteworthy that the character 皿 *śüä* < *\*xiwet* ("the graph is a drawing of a sacrificial vessel with content" per Karlgren, 1940:410 a-c) is used for rendering some Chinese homonymous etyma *sü* (Karlgren quotes four of them sub 410 e and f-h, all of whose prototype he

reconstructs as M.-Ch. \*s<sub>1</sub>juət and A.-Ch. \*s<sub>1</sub>wět 'sollicitude, pity, sorrow', quoted from the Šy-čzing, and 'sollicitude; care about' from the Čžou-li); it is furthermore found as a loan for swət, suət, Modern-Ch. su, 'to rub, brush', all with original s-, not \*x- or \*x'-. The graphemes 410g and h are from Čžou I and Čžou II resp., i.e. from the earliest period of Chinese writing. As to the initial, it would be important to know during which epoch of that period, Čžou I and Čžou II, and in which positions the forms written with 血 originated and for how long a time initial \*x-/x'- and s- or s-/ś- were actually distinct phonemes.

Greek ἵχῶρ 'blood, blood of the gods' which often, however unsuccessfully, was aligned with ἔαρ 'blood', 'αἷμα κύπριου' with Hesychios (cf. Boisacq, 1950:209), together with εἶαρ 'sang, sève, suc' (Boisacq, ib.) would appear to be comparable with A.-Ch. \*xjuet 'blood', the only difficulty being the initial i- in the Greek form. To consider it as a prothesis would be a mere subterfuge, but a nominal composition element might well be hidden under this i-. Persson had thought of \*σι-χῶρ as a prototype of ἵχῶρ (according to Boisacq, 1950:388) where \*σι would go back to the IE root \*sei-/soi-/səi- 'to trickle, etc.', saying, however, nothing about the morphological details. But the other etymon for blood, likewise unexplained, ἔαρ, ἦαρ, εἶαρ seems to remount to an ancient etymological connection with the Uralic etymon for blood, so far known only from the Finno-Ugric languages, not from Samoyedic: Suomi veri, Lapp vâra-/vârrâ-, Mordv. vef, Mari wər, Udmurt, Komi vir, Mañśi ü:r/wü:r, North wigr, Xanty wər, Hung. vér/vără- (Collinder, 1977:137), although it is put to the IE heteroclitic, Skr. ásṛk-, gen. asnāḥ, Old Lat. aser, asser, assyr, Latv. asins, Armen. ariun, Toxar. A jsa:r, and Hittite e-eš-ḫar (ešḫar), gen. ešḫanas (these forms are given in Pokorny, but the use of s and š is not clear), all from an IE \*ēs-ṛ-(ḡ), gen. \*eś-n-és 'blood'; morphologically rather difficult but semantically somehow acceptable, since they all mean 'blood'. That ἔαρ may actually be a cognate of the IE etymon, reconstructed in the above way and historically extant in the forms quoted in Pokorny, 343, seems to be more than doubtful. The Finno-Ugric etymon, Suomi veri 'id.' etc. does so far not appear to have any parallels outside of Uralic. In Greek ἔαρ, etc. no traces of an ancient initial digamma are extant.

Rédei (1986:576) posits PU \*wire, supposing that i in the first syllable in a position before a subsequent r shifted to e, which is otherwise observable in quite a few languages and dialects, parallel to the similar process of o < u, cf. e.g. Middle Rhein-Frankish Wert < Wirt, as in the saying Wer nix wert, wert Wert 'Wer nichts wird, wird Wirt', errt < irrt or korz < kurz, Worscht < Wurst, etc. This supposition of \*wire could be juxtaposed with the above mentioned Dravidian forms. For the Uralic etymon \*kälē 'geronnenes Blut' (Rédei 1986, I:134), a comparison with Old Chinese \*xwe:ð should be considered. The oscillation of the vocalism in the Uralic and Dravidian forms

can be reflected in the Greek variants  $\xi\alpha\rho$ ,  $\eta\alpha\rho$ ,  $\epsilon\lambda\alpha\rho$ . (For the reference to both Uralic words in Rédei (1986) I must thank Irén Hegedűs in Pécs.)

As to Dravidian connections, the following might be considered as possible: 1. Tamil *vaṛi* 'to flow, overflow', Malaja:lam *vaṛijuna* 'id.', Kannaḍa *baḷi* 'to flow out completely', Goṇḍi: *vaṛ-*, *vaṛu:na* 'to soak, drip', Koṇḍa *vaṛ-* 'to drip down (as through filter)', (Burrow and Emeneau, 1986:5296); 2. Tamil *va:ṛ* 'to exist, live, flourish, be happy, etc.', Mal. *va:ṛ* 'life' and numerous derivatives in Koṭa, Toda, Kannaḍa *ba:ṛ*, *ba:ṛu*, *bardunḡu*, etc. 'to live, be alive, subsist, state of living prosperously and happily, etc.', also in Koḍagu, Tuḷu, Telugu, Ko:la:mi:, Najki:, Pa:rṛi:, Goṇḍi:, Kannaḍa, Ku:i:, Kuwi:; in Koḍagu *barykaty* and Tuḷu *barkatu* 'prosperity' a loan via Urdu from Arabic *بركة* *barakatun* 'bliss' is found (Burrow and Emeneau, 1986:5372); 3. Phonetically difficult is: Ta. *vajiru* 'belly, stomach, paunch, womb, center, etc.', Mal. *vajaṛu* 'belly, stomach, inside, receptacle of fruit-seeds, etc.', Koṭa *va:r*, Toda *pa:r* 'belly, pregnancy', Kannaḍa *basaru*, *basir*, 'id., embryo, inside, hold of a ship', Tuḷu *banṛi*, very similar meanings, Koṇḍa *vasiki*, Pengo *vahiṇ*, Maṇḍa *vahiṇ* '(small) intestines', Ku:i: 'intestine, bowels', Kuwi: 'stomach', *wahi* 'intestines', (Burrow and Emeneau, 1986:5259). This etymon, *va:ṛ*, is cognate with Altaic Turkic *ba:r*, *va:r* < *\*ba:r* 'existing, being'. Nowhere is the meaning of 'blood' to be found, which throughout Dravidian is Ta. *nejto:r*, Kannaḍa *nettara*, Telugu *netturu*, *netru*, etc., (Burrow and Emeneau, 1986:3748). Since etyma for 'blood' are often met with in expressions for 'to drip, drizzle, ooze, etc.', some of this latter category may have an etymological connection with the term for 'blood' in Dravidian, as perhaps: 1. Ta. *toṛi* 'to be spilt', Tuḷu *dorijuni* 'to flow, etc.', Goṇḍi: *to:ṛa* 'blood which precedes the birth of a child', *to:ṛg-* '(water) to be spilt', Ku:i: *to:ṛa* 'to be liquid, flow, trickle', (Burrow and Emeneau, 1986:3523); cf. Ta. *čo:r* 'to trickle down as tears, blood, or milk, fall, drop, ooze', *čo:ri* 'blood, rain, shower', *to:rai* 'blood', *to:r* (in Tamil, etc. *nejto:r* 'blood' (Burrow and Emeneau, 1986:3748); *čo:ri* 'to flow down, pour forth, effuse, etc.', *čura* 'to stream out, spring forth, gush, flow', Mal. *čo:ri*, *čo:ra* 'blood', Kannaḍa *suri* 'to flow, pour as tears, rain, blood, etc.', Koḍagu *to:r-* 'to leak (of water, roof, pot)', *čo:ra* 'blood', and many forms and meanings throughout Dravidian (Burrow and Emeneau, 1986:2883). In a few instances semantic transition to 'blood' has taken place. Burrow and Emeneau compare Naha:li: *čorṭo* 'blood', whose relationship with the Dravidian etyma still remains to be investigated.

On Old Latin *assyrr* (sic) and *aser* "blood" cf. the etymological dictionaries by Walde & Hofmann (1938, 1954, I:72) and Ernout & Meillet (1967), s.v. Walde and Hofmann are quite critical of the IE etymology. This term might rather be an ancient Mediterraneo-Caucasian element inherited from pre-Indo-European predecessors of the later Indo-European invaders or peoples of ancient Europe.

Research in this field is, unfortunately, not progressing as expected, because it presupposes a thorough knowledge of Basque, the Etruscan language and their Anatolo-Caucasian cognates, and numerous onomastic and some lexical elements handed down in texts as well as in geographical and person names. As long as the majority of linguists and philologists, with stubborn perseverance in their narrow-gauge solipsistic, often outright agnostic attitude that has its roots in views which had been dominating particularly the humanities in a large part of the Western World during the prelude to the First World War, are unable, often even unwilling to take up seriously comparative research in the only pre-Indo-European language still living on in our days, Basque, as well as Etruscan with quite an amount of well-preserved texts, written in a Greek alphabet, - as long as that majority thus replaces genetic relationship with interference and mixture of languages, typological developments or simply borrowing, this voluminous task has to be carried out by a very small number of linguists, of whom the recipient of this Festschrift, Professor Vitalij Viktorovič Ševoroškin is one, and it is to him that all of us offer our congratulations of his fruitful work.

МЪНОГАЯ ЛЪТА!

### References

- Boisacq, Emile.  
 1954 *Dictionnaire étymologique de la langue grecque* (4th ed.).  
 Heidelberg: Carl Winter.
- Burrow, Thomas and Murray B. Emeneau.  
 1986 *A Dravidian Etymological Dictionary*. Oxford: Oxford  
 University Press. Second edition.
- Castrén, Matthias, A.  
 1858 *Versuch einer Jenisei-Ostjakischen und Kottischen  
 Sprachlehre*. St. Petersburg.  
 1969 *Nordische Reisen und Forschungen*. Leipzig.
- Cincius, V. I. et al.  
 1975-77 *Sravnitel'nyj Slovar' tunguso-man'čžurskix jazykov*, 2 vols.  
 Leningrad: Nauka.
- Cincius, V.I. and L. D. Rišes.  
 1952 *Russko-evenskij slovar'*. Moscow.



- Collinder, Björn.  
1977 *Fenno-Ugric Vocabulary. An Etymological Dictionary of the Uralic Languages.* Second, revised edition, Hamburg.
- Donner, Kai.  
1944 *Kamassisches Wörterbuch nebst Sprachproben und Hauptzügen der Grammatik, bearbeitet und herausgegeben von Aulis J. Joki.* Helsinki.
- Ernout, Alfred and Antoine Meillet.  
1967 *Dictionnaire étymologique de la langue latine. Histoire des mots.* Paris: Klincksieck.
- Hattori, Shirô.  
1964 *An Ainu Dialect Dictionary.* Tokyo.
- Karlgren, Bernhard.  
1940 *Grammata Serica. Script and phonetics in Chinese and Sino-Japanese.* Museum of Far-Eastern Antiquities. Bulletin No. 12. Stockholm.
- Kluge, Friedrich and Walther Mitzka  
1975 *Etymologisches Wörterbuch der deutschen Sprache.* 17th edition. Berlin and New York: Mouton de Gruyter.
- Meyer-Lübke, Wilhelm  
1968 *Romanisches etymologisches Wörterbuch.* Heidelberg.
- Nevskij, A. N.  
1960 *Tangutskaja filologija I-II.* Moscow.
- Pokorny, Julius.  
1959 *Indogermanisches etymologisches Wörterbuch.* Bern: Francke Verlag.
- Räsänen, M.  
1969 *Versuch eines etymologischen Wörterbuchs der Türkssprachen.* Helsinki.
- Rédei, Károly  
1986 *Uralisches Etymologisches Wörterbuch.* Budapest: Akadémiai Kiadó.

Shafer, Robert

1960 "Ethnography in Ancient India."

Shirokogoroff, S.M.

1944 *A Tungus Dictionary*. Tokyo.

Vasilevič, G.M.

1948 *Očerki dialektov evenkijskogo (tungusskogo) jazyka*.  
Leningrad.

1958 *Evenkijsko-russkij slovar'*. Moscow.

Vasmer, Max

1950-59 *Russisches etymologisches Wörterbuch*. Heidelberg.

Wade-Giles = Giles, H. A.

1967 *A Chinese-English Dictionary*. Second edition reprinted  
Taipei.

Walde, Alois. and J.B. Hofmann.

1938, 1956 *Lateinisches etymologisches Wörterbuch*. Heidelberg: Carl  
Winter.

## Altaic Evidence for Clusters in Nostratic

Peter A. Michalove  
*University of Illinois*

1. Illič-Svityč's (1971-84) proposal of a Nostratic language family, encompassing the Afroasiatic, Indo-European, Kartvelian, Uralic, Altaic and Dravidian languages, is still controversial, to say the least, even some 25 years after the publication of the first volume of his posthumous Nostratic dictionary. One reason for the difficulty is that, even if it is correct that some or all of these families are related, or if others are added, even most dedicated adherents of the Nostratic hypothesis agree that many particulars of Illič-Svityč's reconstruction are incomplete or need revision. Thus, we need to distinguish between, on the one hand, problems in the proposal that can be remedied, which would strengthen the case for a genetic relationship between these languages, and on the other hand, flaws so deep that they would lead us to discard entire enterprise.

One of the most difficult problems with Illič-Svityč's reconstruction is the elaborate system of affricates. Starostin (1991:121-122), who accepts the Nostratic hypothesis in general, writes that the affricates "still remain the most problematic area of the Nostratic reconstruction" (my translation, PM).

Specifically, Illič-Svityč proposed an overly rich system of no less than nine affricates. As with the stops, there are three series of affricates, fortis (possibly glottalized) voiceless, lenis voiceless, and voiced. Each of these series of affricates is represented by three articulatory positions, hissing, hissing-hushing, and hushing, represented as follows.

<i>c'</i>	<i>č'</i>	<i>č'</i>
<i>c</i>	<i>č</i>	<i>č</i>
<i>3</i>	<i>3</i>	<i>3</i>

According to Illič-Svityč these affricates were preserved as such in Kartvelian, Uralic, Altaic and Dravidian, while their reflexes are sibilants or stops in Afroasiatic. In Indo-European they correspond to clusters of *s* plus a stop in initial position, with the cluster simplifying medially to *s*.

The development from affricates to clusters, or to simple obstruents, is certainly an unusual phenomenon in those languages of the world whose

histories are well-known, and most textbook treatments suggest that such a development should be rejected as unnatural.

Meillet (1967: 105), speaking of "general formulas of change" writes, "All linguists who have had to examine phonological changes and to establish rules of correspondences between different languages have felt that these changes take place according to certain general types." As an example of a particularly common and natural change, Meillet cites the palatalization of velars in the Romance languages.

Hock (1986:535) explicitly contrasts natural and unnatural directions of development: "Given two otherwise equally acceptable competing analyses, we prefer the one which postulates more natural or common processes." Citing the alternation of Italian *amik-o* 'friend' and its plural *amič-i*, Hock writes, "[W]e could *a priori* reconstruct either [k] or [č] as the root-final consonant and derive the other by means of an appropriate sound change. However, since palatalization is a very common and natural process, while the alternative shift of [+pal] to [+vel.] is not, we will prefer the reconstruction with invariant [k], together with the change...[+vel.] > [+pal.] / \_\_\_\_\_[V, +front]."

Fox (1995:168) makes the same point using the example of Old Church Slavonic *kŭto* 'who' and *čŭto* 'what'. Fox writes, "In the Slavic case, we may appeal to knowledge of likely changes: [k] to [č] is clearly more plausible than the reverse, and we would therefore opt for [k] as the value of the pre-phoneme."

Doerfer (1973) thus objects to Illič-Svityč's proposal, partly on the grounds that the development from affricates to clusters, posited for Indo-European, is an unnatural one, and a "disregard of the empirical data" ("Nichtbeachtung der Empirie"). The same could be said for the development of affricates to stops or hissing sibilants in Afroasiatic. Doerfer cites the apparently unnatural nature of this development as part of his argument that Illič-Svityč's entire construct was false, and that no genetic connection between these families could be maintained.

In addition, if we assume that some of the proto-languages descended from Nostratic inherited affricates, as Illič-Svityč claims, then we face a similar problem of naturalness in accounting for cases, as we will see below, in which some of the modern attested languages have a simple stop corresponding to a form that Illič-Svityč proposed as an affricate.

2. Manaster Ramer (1994) responds to Doerfer's argument by suggesting that this weakness, and numerous others in Illič-Svityč's reconstruction, do not in themselves invalidate the possibility that the Nostratic theory is correct. Rather, he proposes that the weaknesses can be remedied and thereby actually strengthen the case for Nostratic.

Specifically, Manaster Ramer suggests that the phonemes Illič-Svityč posited as affricates were originally clusters of a sibilant plus stop. This reconstruction would mean that the clusters of Indo-European are archaic, and the affricates found, e. g. in Kartvelian and elsewhere, are secondary developments. If this

analysis is correct, then the development from clusters in Nostratic to affricates in the non-IE branches would be a natural and commonly attested one, answering an important criticism of the proposed Nostratic system. In addition, a series of clusters in Nostratic satisfies the criticism of an excessively rich affricate system.<sup>1</sup>

In fact, Indo-European and Kartvelian are not the only families associated with Nostratic that show evidence of inherited clusters instead of affricates. For Afroasiatic, the sibilant and stop reflexes that Illič-Svityč proposed are much easier to imagine developing from clusters of *s* plus stop than from affricates, where there would be similar problems of naturalness.

In Uralic, Collinder's (1960) reconstruction posits two affricates, *č* and *ć*, of which the hushing form, *č*, yields obstruent reflexes in most of Samoyed, while *ć* has obstruent reflexes in Motor and Taigi as well as some Saami dialects (Collinder, 1960, Rédei, 1986). Janhunen (1981a, 1981b) and Sammallahti (1988) reconstruct a somewhat different system, using only one affricate, but the naturalness of the non-affricate reflexes in some Fenno-Ugric dialects and Samoyed (Mikola, 1988) still remains to be explained.

To my knowledge no one has ever questioned the genetic status of the Uralic family on the grounds that it requires us to assume an unnatural phonological development to arrive at some of the daughter languages. However, if we suppose that Uralic inherited clusters of a sibilant plus stop, and that the clusters developed to affricates in most of the Fenno-Ugric languages, and simplified to single stops in others, we solve the problem of naturalness within Uralic, and we arrive at a situation consistent with Indo-European and Kartvelian.<sup>2</sup>

Without going into further detail about Afroasiatic and Uralic, however,

---

<sup>1</sup> For the record, Manaster Ramer (1994) also cites Mačavariani's (1960) and Klimov's (1964) reconstruction of Kartvelian, in which proto-Kartvelian affricates developed to clusters in all the daughter languages except Georgian. If Mačavariani's and Klimov's reconstruction is correct, this would provide a precedent for the apparently "unnatural" development to clusters in Nostratic. However, Manaster Ramer clearly prefers Schmidt's (1961, 1962) reconstruction, in which the clusters of Laz, Mingrelian, and Svan are original, and the affricates of Georgian are secondary. More recently, Teselec (1995) proposes a similar analysis. Schmidt's and Teselec's formulation is in line with most scholars' views of naturalness, and Manaster Ramer takes it as a model for the development from Nostratic to Kartvelian and the other daughter branches with attested affricates.

<sup>2</sup> Décsy (1990:28) addresses this problem by reconstructing a palatalized *\*tj* in place of *\*ć*. He writes, "It is a universal that a *c* can stem from a *t* but not vice versa, i.e. from a diachronic point of view, *t* is always older than *c*." Décsy's solution is certainly possible. The solution of clusters, however, is equally plausible from the point of view of Uralic alone, and in addition, it is consistent with what we know of the extra-Uralic situation.

the focus of this paper will be to consider the evidence that Altaic also inherited clusters rather than the traditionally reconstructed affricates.<sup>3</sup>

3. Poppe (1960) posits two affricates for Proto-Altaic, \*č and \*č̣. Starostin (1991) writes Poppe's \*č as \*č', and posits a separate \*č̣, based on a new set of correspondences, while Starostin's \*č̣ corresponds to the same voiced affricate as in Poppe. Thus, Starostin arrives at three affricates in Altaic, corresponding to Illič-Svityč's three series, fortis voiceless, lenis voiceless, and voiced. Within each of the series, Starostin (and Illič-Svityč) propose that the three articulatory positions merged so that, for example, Nostratic \*c', \*č', and \*č̣' yielded Altaic \*č'. Illič-Svityč proposes a similar merger to clusters in Indo-European, so that the three fortis voiceless Nostratic affricates are realized as IE \*sk (with a subsequent split to \*sk, \*sk' and \*sk<sup>w</sup> resulting from the quality of the following vowel). Therefore, our primary concern here is to consider the evidence for inherited clusters in Altaic, and their relation to Indo-European forms, which would appear to preserve the original clusters most clearly.

To begin with Altaic \*č', Starostin lists the reflexes of this phoneme as č in Turkic, Mongolian, Tungusic and Korean, but *t* in Japanese. For example, PA \*č'ä:k'V<sup>4</sup> 'time' yields PT \*ča:k, OT ča:q 'gerade, genau' (v. Gabain, 1977), Turkish çağ 'time, era, period'; WM čaγ 'time'; MK čäk; and PJ \*təki, OJ toki. It would be difficult and unnatural to derive the Indo-European clusters from inherited affricates, as Illič-Svityč attempted to do. But if we suppose that PIE and Altaic inherited not \*č', but the cluster \*sT' (where T' is a cover symbol for the Nostratic fortis voiceless stops; these correspond to the IE voiceless stops, which we may represent as T), then we are positing a simple and well-attested development. The cluster could easily simplify to the affricate in PA, yielding affricates in the western Altaic languages, and simplifying to a single obstruent in Japanese.<sup>5</sup>

<sup>3</sup> Altaic, of course, is at least as controversial in its own right as is Nostratic, and I am not sure that rearguing the case for Altaic here is going to change any minds. For purposes of this paper I treat the Turkic, Mongolian, Tungusic, Korean and Japanese languages as a genetic family, based primarily on the treatment in Starostin (1991).

<sup>4</sup> Altaic reconstructed and attested forms are per Starostin (1991), except where otherwise indicated.

<sup>5</sup> Alexander Vovin (personal correspondence) states that the Altaic affricates are realized as palatal stops in Tungusic, in the more conservative northern Korean dialects, and possibly in some Turkic and Mongolic dialects. Therefore, Vovin reconstructs a series of palatal stops for PA, rather than an affricate series, and he sees the affricates in most of the western Altaic languages as a secondary development from the palatals. He also considers the non-palatal stops of Japanese (and within

The Altaic form \*č'epV- 'twist, wind' is reflected in PT \*čebir- 'twist, turn', OT čevür- 'id.'; PTng \*č/e/b- 'twist, roll, E čiwär- 'id.'; and Japanese *tawam*- 'bend, twist' (Miller, 1971: 85). If we reconstruct an initial cluster here as \*sT'epV- in place of \*č-, then the form corresponds phonologically to PIE \*(s)kerb(h) 'turn, curve'. The development, and hence the relationship itself, now become easier to accept.

If we are to equate these forms, we must also account for the -r- in PIE \*(s)kerb(h), or rather its loss in Altaic. We will discuss this shortly, as well as the significance of the mobile-s in the PIE form.

A similar correspondence is PA \*č'arV- 'cut, tear apart' : PIE (s)ker 'cut.' Here the Altaic form is reflected in Tungusic (Evenki) čari- 'tear'; and PJ \*tät, OJ tat- 'cut' (although we should note that the Evenki form is isolated within Tungusic).

There are several similar examples in Illič-Svityč (1971), which did not consider Japanese, so we do not have clear non-affricate reflexes in these. However, the correspondence of PJ \*t to \*č in the other Altaic languages is amply documented in Starostin (1991). The first two examples below are from Illič-Svityč (1971-84), and the third is from Starostin (1991).

PA \*č'ap(a)- 'chop, beat' : PIE\*(s)kep- 'cut, split (with a sharp tool)'. OT čap- 'beat'; WM čabči- 'chop'; Nanai čapči- 'chop', Udehe čabča- 'chop wood', eastern E čapka 'fish-spear, harpoon', Negidal čapkala:- 'chop with a fish-spear' (Cincius, 1975-77, v. 2: 384).

---

Tungusic, of Orok) as secondary developments resulting from the loss of the palatal feature in these languages.

This is certainly a plausible sequence of events, and for the study of Altaic alone, perhaps we do not need to pursue the matter further. But to consider the possibility of an Altaic affinity with neighboring languages, we must wonder about the origin of the affricates (or palatals), which typically arise from some combination of simpler elements, possibly in a conditioned environment, and often with differing reflexes in the various daughter languages. For example, the various reflexes of the Altaic affricates are similar to the varied Slavic reflexes of inherited jotted dentals or \*kt-, or of Common Slavic \*t before a front vowel. Thus, the reflexes of the Altaic affricates are consistent with a pre-Altaic source in a cluster and, as we will see, this possibility provides further clarification of extra-Altaic comparisons as well.

In addition, Karl Heinrich Menges (personal correspondence) states that the Tungusic and other Altaic forms in question are indeed affricates, and not palatal stops. If Menges' view is correct, this would remove Vovin's obstacle entirely. I have not personally heard native speakers of any Tungusic languages.

PA \*č'alu- 'split, cut' : PIE \*(s)kel- 'cut'. OT čal- 'beat'; WM čali 'edge'; Nanai ča:li- 'cut off', Ulcha čalu- 'cut (off)', E čali: 'point of an arrow'; Korean čari 'cut off'.

PA \*č'akV- 'white' : PIE \*speng- 'glitter'.  
PT \*čakir 'bright grey', OT čaqir 'id.'; PM \*čaga-γan 'white.', MM čaqa:n 'id.'; PTng \*čag- 'id.', Ulcha ča:gža(n) 'id.'

This reanalysis forces us to reconsider the root structure of Nostratic and Proto-Altaic. Both are traditionally reconstructed as CV(C)CV. If we suppose that the Altaic affricates represent inherited clusters, then the structure of both can now be stated as (C)CV(C)CV. This is, of course, consistent with Indo-European root structure, except for the final vowel, which is not typical of IE.

However, within a structure (C)CV(C)CV, Altaic imposed a further constraint. Starostin (1991) lists about thirty Altaic roots with medial clusters, retained most clearly in Tungusic, but none of these have the initial affricates that we have been concerned with here. Therefore it appears that proto-Altaic allowed initial clusters and medial clusters, but not both in the same form. In cases with Nostratic initial and medial clusters, Altaic first simplified the medial cluster by dropping the non-obstruent form.

Thus, from our examples above, Nostratic forms such as \*sk'erpV- and \*sp'ankV- simplified the medial clusters to Altaic \*sk'epV- and \*sp'akV-. Then, the simplification of initial clusters to affricates would have proceeded as we have discussed to \*č'epV- and \*č'akV-. This produced the situation we are familiar with, in which Altaic does not permit initial clusters; medial clusters remained in forms that did not reflect an inherited initial cluster.

From the Indo-European side, we have seen that the mobile *s-* is very common in these sets. While various explanations have been offered for the mobile *s-*, the examples here suggest that the sibilant element was original, and that IE was subject to early pressure to simplify clusters of this type, although not in as thoroughgoing a manner as we see in Altaic and some of the other branches of Nostratic.

4. The Altaic lenis voiceless affricate, \*č, was not in Illič-Svityč's system, but Starostin (1991) reconstructs it on the basis of forms in which Turkic and Mongolian \*d correspond to Tungusic \*ž; PK \*č; and (in initial position) to PJ \*t. An example is Altaic \*čawVr'V 'salt, bitter, sour,' reflected in PT \*du:r' 'salt', OT tuz 'id.', Tm. du:z 'id.'; PM \*dabu-sun (<\*dabur-sun) 'id.', MM dabu-sun 'id.'; PTng \*žujar- 'bitter, sour', Nanai žojor-si 'id.', E žuin/žujur 'bitter'; PK \*čjə:r-, South Korean če:l- (<čeber-) 'become salty'; PJ \*túra-, Old Japanese tura 'bitter, heavy, unbearable.'



Here the Turkic and Mongolian reflexes, as well as the Japanese, yield a non-palatal dental stop, while Tungusic and PK give affricates (phonetically, palatals). Following our earlier reasoning, if Altaic \*č', developed from a Nostratic \*sT' (which yielded IE \*sT), then we may suppose that the Altaic \*č developed from Nostratic \*sT, which we would expect to correspond to IE \*sD (where D represents any voiceless stop in IE).

For its part, Indo-European underwent a subsequent simplification of the cluster \*sD to \*s. Starostin lists the first four examples below in which Altaic \*č corresponds to IE \*s, and we add the fifth.

PA \*čawVr'V 'salt, bitter, sour (see above) : PIE \*sH<sub>3</sub>ew-ro 'id.'

PA \*čualV 'type of leafy plant': PIE \*(s)wVly-k 'willow'.

Altaic shows the reflex č in PTng \*žali-*kta* 'leafy plant.', E žali-*kta* 'id.'; PK \*čür-*ki* 'stalk, branch without leaves', MK čür-*ki* 'id.', but obstruent reflexes in PT \*dal 'willow palm tree, branch, OT tal 'twig' Tm. tal 'willow', Salar da:l 'wood'; PM \*dolu-*γana* 'hawthorn', MM dolu-*gono* 'id.'

The Indo-European form is of particular interest here. Friedrich (1970) compares two forms, a western IE \*salyk, on the strength of Celtic, Italic, Germanic and Greek; and a Central IE form with initial \*w-, attested in Germanic, Greek and possibly Hittite.

Friedrich combines these in the alternation \*sVlyk- ~ \*wVlyk-, which he writes \*s/wVlyk-. Of course, this is the familiar mobile s-, as Friedrich notes, citing Meillet (1937: 171-72), who gives examples of the mobile s- before stops and sonorants. Friedrich suggests that the form "may also represent a k-extension added to an old i-stem," and the Altaic form is consistent with this supposition. The result is our form given above, \*(s)wVly-k.

PA \*ča:(w)tu 'sweet, pleasant to taste': PIE \*sueH<sub>2</sub>d- 'sweet'.

Altaic affricates in PTng \*žuti: 'delicious', E žuti: 'id.', Even žut 'sweet' (Cincius, 1975, v. 1); but obstruent reflexes in PT \*da:t 'taste to test, have (pleasant) taste, OT tat-, Brahmi texts ta:tt- 'schmecken' (v. Gabain, 1974), Turkish tat-lı 'sweet', Tm. da:t-lı 'delicious' (Baskakov et al., 1968); WM dadu- 'become accustomed to.'

PA \*čür'ü- 'to string, put in line': PIE \*ser- 'order consecutively, tie'.

Altaic affricates in MK čiri-tá 'straight ahead, go straight', čür-hjè-tá 'to string, order', and obstruents in PT \*dir', dür'; 'straight, even, OT tüz 'ordnen, eben machen', Turkish düz 'id.' (v. Gabain, 1974); WM dürü 'stick in, push in; PJ tura 'row, string' (Unger, 1993: 152). See Starostin (1991: 13, 14) for his treatment of the Altaic forms as two separate etymologies.

PA \*č<sub>o</sub>(:)/lu- 'full, to fill' : PIE \*sH<sub>3</sub>elwo- 'well kept, whole'.

Altaic affricates in PTng \*ž<sub>alu</sub>(-m) 'id.', Negidal ž<sub>alum</sub> 'id.'; PK \*č<sub>är</sub>à 'be sufficient', MK č<sub>är</sub>à 'id.', and obstruents in PT \*do:l-i 'full', OT *tolu* 'id.', Tm. *do:l-i* 'id.'; PJ \*t<sub>är</sub>- 'be sufficient, full, seize', OJ *tar*- 'id.';

5. The Altaic voiced affricate, \*ž, represents forms with Turkic \*y; Mongolian and Tungusic \*ž; Korean \*č; and Japanese \*d. An example is PA \*ž<sub>ian</sub>V- ~ ž<sub>ain</sub>V- 'burn, ashes, tar', with PT \*yan- 'burn, catch fire', Turkish, Tm. *yan*- 'id.'; PTng \*ž<sub>ian</sub>- 'to flame', E ž<sub>a:n</sub>je- 'id.'; PK \*č<sub>äi</sub> 'ashes', MK č<sub>äi</sub> 'id.'; PJ \*d<sub>äni</sub> 'tar', OJ *yani* 'id.' Here again, it is the non-palatal obstruent, *d*, in PJ that speaks most clearly against an original affricate.

Of course, the OJ reflex, *y* is not a stop, and the PJ reconstruction of \**d* is based largely on the Yonaguni dialects that have *d* corresponding to OJ *y*. Martin (1987) and Starostin (1991) consider the possibility that the Yonaguni *d* is a secondary development, but Martin rejects that option on the same grounds we have been using in this paper: a "more natural hypothesis would have the main stream dialects lenite earlier stops." (Martin 1987:20)<sup>6</sup>.

Thus, following our reasoning so far, we posit a source for the Altaic form in the Nostratic cluster consisting of \**s* plus a voiced stop, \**sD*, which would yield IE \**sDh*. As with the previous set, the corresponding IE clusters were simplified to \**s*, as in the following forms.

PA \*ž<sub>ian</sub>V- ~ ž<sub>ain</sub>V- 'burn, ashes, tar' (see above) : PIE \*senk- 'to burn, dry'.

\*ž<sub>e</sub>- 'eat', \*ž<sub>o</sub>ž(V)-*k'a* 'fat, abundant' : PIE \*seH<sub>2</sub>-, \*seH<sub>2</sub>t 'sated.' Starostin (1991) reconstructs the Altaic forms as two separate roots. However, the second appears to be a derived form with expressive reduplication, and we will treat them together here.

Altaic has the affricate ž in PM \*ž<sub>u</sub>ž<sub>a</sub>ŋa-n 'fat', MM ž<sub>u</sub>ž<sub>a</sub>ŋan 'id.'; PTng \*ž<sub>e</sub>-*p*- 'eat', E ž<sub>e</sub>-*p*-, ž<sub>e</sub>-*b*- 'id.'; PK \*č<sub>a</sub>- 'eat', MK č<sub>a</sub>-*si*- 'id.'; and the obstruent \**d* in PJ \*da-pa 'hungry', \*d<sub>ütä</sub>-*ka* 'huge' OJ *yapa* 'hungry', *yuta*-*ka* 'huge'. Turkic also has *y* in PT \*ye:- 'eat', \*yogan 'fat', OT *ye*- 'eat', *yo*ŋun 'fat'.

PA \*ž<sub>uab</sub>V 'weak, exhausted, poor' : PIE \*swep- 'sleep'.

<sup>6</sup> Martin (1987: 20) states that PJ "d was surely palatalized and possibly affricated, [d<sub>y</sub>] or [dž]". This suggests that OJ \*d passed through a stage similar to attested Mongolian and Tungusic before developing to non-Yonaguni *y*. If so, we may surmise that Turkic followed a similar development.

Altaic shows *ž* in WM *žoba-* 'suffer, worry'; PTng *\*žowa-* 'become poor, suffer'; E *žoyo-* 'be poor, need, trouble oneself' (Cincius 1975-77, v. 1: 260-61); Japanese has *\*d* in PJ *\*duawa* or *\*dauwa* 'weak', OJ *yuowa-* 'id.'; Turkic also has *y* in PT *\*yabri-* 'to tire', *yab-ir* 'exhausted, weak', OT *yawri-* 'elend, schwach werden' (v. Gabain, 1974).

PA *\*žarV-* 'send' : PIE *\*se:(i)-* 'dispatch'.

PM *\*žaru-* 'send, assign a task', MM *žaru-* 'id.' WM *žaru-da-su(n)* 'slave, servant, messenger'; PJ *\*dára* 'send away, dispatch' (Unger, 1993: 144), OJ *yar* 'id.'

The following forms appear to have no Japanese cognates, but illustrate the correspondence PA *\*ž* to PIE *\*s*:

PA *\*žilV* 'slip, slippery, smooth' : PIE *\*(s)leib* 'slippery, to slide'.

PM *\*žili-*, *žilu-* 'smooth, even'; MM *žilim*, *žilum* 'id.'; PTng *\*žulV* 'smooth', E *žula:-kin* 'bare'; PT *\*yilan* 'snake', OT *yılan* 'id.' The Turkic semantic shift is somewhat oblique, but the phonological correspondences are precisely what we would expect.

PA *\*žur'u-* 'float, current' : PIE *\*srew-* 'flow'.

PT *\*yür-*, OT *yüz-* 'id.'; PTng *\*žurku* 'fast current', E *žurku* 'id.' (Cincius, 1975-77, v. 1: 277), Negidal *žojku* 'sea-channel'.

These examples show several cases in which Altaic inherited a non-initial cluster in addition to the initial form we have been concerned with. As with the forms in *\*č*, Altaic first underwent a development in which the non-initial clusters were simplified, but in these voiced cases the process was somewhat different than the one we saw with *\*č*.

Here, the simplification entailed two steps. First, clusters involving a liquid were simplified by metathesis, as Nostratic *\*sDr'uV-* > *\*sDur'V-* > Altaic *\*žur'V-*; and Nostratic *\*sDilbV* > Altaic *\*sDilbV*.

This last form still leaves us with an initial and a root-final cluster, however. This was resolved in a second step, in which the obstruent term of remaining non-initial clusters was lost, e.g., *\*sDilbV* above > *\*sDilV* > Altaic *\*žilV*; and Nostratic *\*sDiank* > *\*sDian* > Altaic *\*žian-*. After these non-initial clusters were simplified in forms with initial clusters, the initial clusters developed to affricates as we have discussed, again yielding the situation we are familiar with in the attested languages. Non-initial clusters are allowed only in forms that do not reflect an inherited initial cluster, and initial clusters are replaced in all cases by palatals.

An important point, which increases our confidence in the system, is that the correspondences offered so far provide us the opportunity to reject otherwise promising parallels in which the phonology does not work. Just as we do not consider Latin *deus* and Greek  $\theta\epsilon\acute{o}\varsigma$  or Latin *dīēs* and English *day* to be cognate because the correspondences do not hold, so, we can apply the same standards in Nostratic and Altaic or Indo-European.

Thus, we expect Altaic \*č' to correspond to IE \*(s)T rather than to \*s, so we would reject a relationship between Altaic \*č'ok'ü- 'to lean, sink, die' with IE \*seng<sup>w</sup>- 'to fall, sink.' In this case, the correspondence of Altaic \*k' to IE \*g<sup>w</sup> also indicates that the similarity is spurious. Similarly, since we expect Altaic \*č and \*ž to correspond to IE \*s rather than to \*(s)T, we can reject parallels like Altaic \*čurV- 'stand' with IE \*steH<sub>2</sub>- 'id', and Altaic \*žä:jV- 'arrow, sharp' with IE \*ske(:)i- 'shoot, throw, hunt.'

6. Thus, of the six families treated by Illič-Svityč, the phonemes he reconstructed as affricates are reflected as clusters in IE, while Afro-Asiatic has sibilants and stops. As we have seen, there is evidence in Altaic, Uralic, and Kartvelian that the attested affricates reflect inherited clusters.

Only Dravidian consistently has affricates and, if we accept the relationship of these six families, the weight of the evidence would now suggest that the Dravidian affricates and the affricates attested elsewhere have their origin in clusters of a sibilant plus stop, a natural and well-attested phonological development. In the case of Dravidian, the development from clusters to affricates is consistent with the structure of Proto-Dravidian. Proto-Dravidian did not allow word-initial clusters, and so would have undergone some form of simplification here. The proto-language is reconstructed with no fricative phoneme (Zvelebil, 1990), so the affricate \*c is a convincing development.

This reanalysis of the source of Altaic and other affricates refines Illič-Svityč's theory and strengthens it in at least three ways: it avoids recourse to unnatural routes of phonological development; it replaces Illič-Svityč's overly rich system of affricates with a more credible system of consonant clusters; and it gives us a criterion to identify superficially similar forms that we can reject as spurious matches.

In all it seems more constructive to address weaknesses of the theory and work to remedy them, as we have tried to do here, than to dismiss the entire theory because of isolated weaknesses. The refinements discussed here, and others that have been proposed since Illič-Svityč's premature death are reasons for the wider linguistic community to give the Nostratic theory a fuller hearing than it has received so far.

Alexis Manaster Ramer, Karl Heinrich Menges, and Alexander Vovin contributed helpful and challenging suggestions to this paper, for which I am very grateful. Of course, they are not responsible for the conclusions stated here.

## Abbreviations

E.	Evenki	PJ.	Proto-Japanese
J.	Japanese	PK.	Proto-Korean
MK.	Middle Korean	PM.	Proto-Mongolian
OJ.	Old Japanese	PT.	Proto-Turkic
OT.	Old Turkic	PTng.	Proto-Tungusic
PA.	Proto-Altaic	Tm.	Turkmen
PIE.	Proto-Indo-European	WM.	Written Mongolian

## References

- Baskakov, N. A., B. A. Karryev, M. Ja. Xamzaev, eds.  
 1968 *Turkmensko-russkii Slovar'*. Moscow: Izdatel'stvo "Sovetskii Enciklopedija."
- Cincius, V. I.  
 1949 *Sravnitel'naja Fonetika tunguso-man'čžurskix jazykov*. Leningrad: Učpedgiz.  
 1975-77 *Sravnitel'nij Slovar' tunguso-man'čžurskix jazykov*, 2 vols. Leningrad: Nauka.
- Clauson, Gerard  
 1981 *An Etymological Dictionary of Pre-thirteenth-century Turkish*, 2 vols. Szeged: Universitas Szegediensis de Attila Jozsef Nominata.
- Collinder, Björn  
 1960 *Comparative Grammar of the Uralic Languages*. Stockholm: Almqvist & Wiksell.
- Décsey, Gyula  
 1990 *The Uralic Protolanguage: A Comprehensive Reconstruction*. Bloomington: Eurolingua.
- Doerfer, Gerhard  
 1973 *Lautgesetz und Zufall: Betrachtungen zum Omnikomparativismus*. Innsbruck: Sprachwissenschaftliches Institut der Universität.

- Fox, Anthony  
1995 *Linguistic Reconstruction: An Introduction to Theory and Method.* Oxford: Oxford University Press.
- Friedrich, Paul  
1970 *Proto-Indo-European Trees.* Chicago: University of Chicago Press.
- von Gabain, Annemarie  
1974 *Alttürkische Grammatik.* 3rd Edition. Wiesbaden: Otto Harrassowitz.
- Hock, Hans Henrich  
1986 *Principles of Historical Linguistics.* Berlin: Mouton de Gruyter.
- Illič-Svityč, V. M.  
1971-84 *Opyt sravnenija nostratičeskix jazykov.* Moscow: Nauka.
- Janhunen, Juha  
1981a "On the structure of Proto-Uralic." *Finnisch-Ugrische Forschungen.* 44:23-42.  
1981b "Uralilaisen kantakielen sanastoa. *Journal de la Société Finno-Ougrienne* 77: 219-74.
- Klimov, G. A.  
1964 *Etimologičeskij slovar' kartvel'skix jazykov.* Moscow: Akademia Nauk SSSR.
- Mačavariani, G. I.  
1960 "O trëx rjadax sibiljantyx spirantov i affrikat v kartvel'skix jazykax." *XXV meždunarodnyj kongress vostokovedov: Doklady delegacii SSSR.* Moscow: Izd. vostočnoj literatury.
- Manaster Ramer, Alexis  
1994 "Clusters or Affricates in Kartvelian and Nostratic" *Diachronica* XI, 2: 157-70.
- Martin, Samuel E.  
1987 *The Japanese Language Through Time.* New Haven: Yale University Press.

- Meillet, Antoine  
 1964 *Introduction à l'étude comparative des langues indo-européennes*. University, Alabama: Alabama Press.  
 1967 *The Comparative Method in Historical Linguistics*. Translated by Gordon B. Ford, Jr. Paris: Librairie Honoré Champion.
- Mikola, Tibor  
 1988 "Geschichte der samojedischen Sprachen." in Sinor, 1988.
- Miller, Roy Andrew  
 1971 *Japanese and the Other Altaic Languages*. Chicago: University of Chicago Press.
- Pokorny, Julius  
 1959 *Indogermanisches etymologisches Wörterbuch*. Bern: Franke.
- Poppe, Nikolaus  
 1960 *Vergleichende Grammatik der altaischen Sprachen, Teil I: Lautlehre*. Wiesbaden: Otto Harrassowitz.
- Rédei, Károly  
 1986-88 *Uralisches etymologisches Wörterbuch*. Budapest: Akadémiai Kiadó.
- Sammallahti, Pekka  
 1988 "Historical phonology of the Uralic languages." in Sinor, 1988: 478-554.
- Schmidt, Karl Horst  
 1961 "Sibilanten und Affrikaten-korrespondenzen in den Kartwel-sprachen." *Bedi Kartlisa* 11/12:149-73.  
 1962 *Studien zur Rekonstruktion des Lautstandes der süd-kaukasischen Grundsprache*. Wiesbaden: Franz Steiner.
- Sinor, Denis, ed.  
 1988 *The Uralic Languages: Description, History and Foreign Influences*. Leiden: E. J. Brill.
- Starostin, S. A.  
 1991 *Altajskaja Problema i proiskhoždenie japonskogo jazyka*. Moscow: Nauka.
- Teselec, Ya. G.  
 1995 "Sibilanty ili komplekсы v prakartvel'skom?" *Voprosy jazykoznaniya*, 1995/2: 10-28.

Unger, J. Marshall

1993

*Studies in Early Japanese Morphophonemics*, 2nd edition.  
Bloomington: Indiana University Linguistics Club.

Zvelebil, Kamil

1990

*Dravidian Linguistics: An Introduction*. Pondicherry:  
Pondicherry Institute of Linguistics and Culture.



## New Albanian Etymologies (Balkan Etymologies 116-145)\*

Vladimir Orel  
Tel Aviv University

This paper, written in order to express my respect and admiration for Vitaly Shevoroshkin, results from a continuous work on the new *Albanian Etymological Dictionary* (Orel 1996) due to be finished next year. The analysis of a vast corpus of lexical material inevitably leads to new etymological versions and solutions some of which are presented below in a fairly condensed form dictated by the genre of a reference book and, more or less, as they will appear in Orel (1996).

**afër** adv., prep. 'near'

From PALb (= Proto-Albanian) *\*apsera* continuing IE *\*apero-*, a derivative of *\*apo-* (Skt *ápara-* 'posterior, later', Goth *afar* 'after' and the like), and reflecting traces of a secondary influence of *\*aps*, a variant of IE *\*apo* reflected by Gk *ἄψ* 'backwards'. Possible but much less probable is the borrowing of *afër* from Germanic: Goth *afar*, OHG *avar* 'again' and other similar forms. ◇ Meyer 1891:3 (borrowed from Lat *\*affināre* 'to approach' ~ *affinis* 'near' with the Gheg form borrowed from Tosk); Jokl 1911:103 - 104 (preposition *a* followed by *-fër* borrowed from Goth *fera* 'side'); Barić 1954:87 (links *afër* to Lat *spernō* 'to sever, to separate, to remove', Gk *σπαίρω* 'to gasp, to pant, to quiver'), 1955:71; Tagliavini 1937:67; Frisk 1960:204; Pokorny 1959:53-54; Mayrhofer 1953:38; Çabej 1976a: 28 - 29 (privative *a-* < *\*n-* and *-fër* compared with Eng *far*); Huld 1984:36.

**ajkë** f 'cream, wool fat'

In dialects, a more phonetically archaic form *alkë* has been preserved. Goes back to PALb *\*alkā* related to Lith *álkti* 'be hungry', *alka* 'hunger', Slav *\*olkti* 'be hungry'. ◇ Meyer 1891:5 (from Lat *alica* 'kind of grain, spelt' with an obvious discrepancy of meaning); Fraenkel 1962:8; Çabej 1976a:31 - 32 (reconstructs *\*olka* and compares *ajkë* with Lat *alga* 'sea-weed'?!).

---

\* For the preceding article of this series see *Etimologija* 1986-1987. Moscow, Nauka, 1989, 220-227.

**ashkë** f 'wood splinter'.

From PALb *\*akskā*, a derivative of IE *\*aks-* 'axis': Skt *ákṣa-* FIND, Gk ἄξων, Lat *axis* and the like. ◇ Meyer 1891:17 (borrowed from Lat *\*ascla*); Jokl 1923:104 - 105 (supports Meyer); Frisk 1960:116.

**beronjë** f 'barren woman; holly; kind of serpent'

Another phonetic variant is *buronjë*. A derivative with a feminine suffix *-onjë* of the unattested *\*ber* < PALb *\*bara* 'naked, barren', related to OHG *bar* 'bare', ON *berr* id. ◇ Meyer 1891:33 (comparison with *berr* and Slav *\*baranъ* 'ram').

**bërcel** m 'kind of wheat, Triticum monococcum'

Derived from the unattested *\*bërcë* ~ *\*bricë* borrowed from Slav *\*bъrica* > Bulg *brica* 'kind of white wheat'. ◇ Jokl *apud* Çabej 1976a:62 (related to *bardhë*); Trubachev 1976:125.

**bisk** m 'branch, twig'

Borrowed from a diminutive Slav *\*bičьkъ* derived from *\*bičь* 'whip', with *-s-* continuing PALb *\*-tš-*. As to *bisk* 'rivulet', it may also belong here. ◇ Meyer 1891:37 (from NGk βίτσα 'switch, rod' borrowed from Bulg *vica* id.); Trubachev 1975:94.

**bosht** m 'spindle, axis, axle'.

From PALb *\*bāsta* close to Germ *\*bōsta* > OHG *buost* 'rope made of bast'. Further related to Germ *\*basta* 'bast' as well as Lat *fascis*, Alb *bashkë*. The spindle is, thus, described as 'juncture'. Note that *boshtër* 'Forsythia' is derived from *bosht* (Çabej 1976a:75). ◇ Meyer 1891:42 (derived from Ital *bosso*); Tagliavini 1937:86.

**burg** m 'prison, stable'

Borrowed from Germ *\*burgaz* 'borough, fenced area' (Goth *baurgs*, OHG *burg* and the like). ◇ Miklosich 1871:7 (from VLat *\*burgus*); Meyer 1891:54 - 55 (various untenable guesses).

**cermë** f 'arthritis'

Borrowed from Slav *\*čьrmъ* 'inflammation' attested in South Slavic only as Slovene *črm*. ◇ Trubachev 1977:149; Çabej 1976a:90 (historically identical with *thermë* - this view can be only accepted for *cërmë* 'cramp, spasm').

**çandër** f 'prop, support'

From \**stšentra* reflecting a singularized plural of the Indo-European neut. \**skentrom* with *s*-mobile, close to IE \**kentrom*: Gk κέντρον 'goad, spur', cf. also Latv *sīts* 'spear, lance' < Balt \**šintas*. The anlaut *ç(a)*- excludes the possibility of a borrowing from Latin or a Romance language, cf. *qendër* 'center'. ◇ Pokorny 1959:567; Frisk 1960:820 - 821.

**çerr** m 'wren'

A substantivized use of a borrowed Slavic adjective \**čьrnъ* 'black'. ◇ Trubachev 1977:155 - 157.

**dangë** f 'belly'

Another variant is *dëngë*. Goes back to PALb \**dangā* etymologically identical with Lith *dangà* 'table-cloth, cover', Latv *danga* 'puddle, marshland', Slav \**doga* 'arc'. All these forms are deverbatives related to Lith *dengiù*, *deñgti* 'to cover'. Adjectival *dëng* 'full, stuffed up' continues PALb \**danga* and also belongs here. As to *deng* 'bundle, full sack', it is rather a borrowing from Turk *denk* 'bale' (Meyer 1891:63) than a cognate of the above forms. ◇ Meyer 1891:61 (to Slovene *danka* 'rectum'); Fraenkel 1962:88 - 89; Çabej 1976a:106 (to *deng*), 121; Trubachev 1978:98 - 99.

**dëllinjë** f 'juniper'

A more archaic variant *dëllënjë* seems to reflect PALb \**daislanjā* (for the derivational structure cf. *mëllënjë*) related to *dell* 'sinew' < \**daislā*. Semantically, the juniper is described as a wiry, sinewy plant, cf. Russ *vt;tdtklybr* id. derived from Slav \**mozgъ* 'brain, marrow', Lith *māzgas* 'knot'. ◇ Meyer 1891:65 (from Lat \**cedrulanea* or \**cedrulina* derived from *cedrus* 'cedar, juniper'); Vasmer 1921:9 - 10 (to Lith *dūlis* 'fog', Skt *dhūli*- 'dust' and the like), 1970:637; Jokl 1923:191 - 193 (same as Vasmer); Fraenkel 1962:426 - 427; Çabej 1976a:121 (related to *daltë* and *dalloj*).

**dërgoj** 'to send'

Borrowed from Lat *dēlēgāre* id. with an irregular change of *liquida*. ◇ Camarda 1864a:67 (to Gk τρέχω 'to run'); Meyer 1891:65 (borrowing from Lat *dirigere* 'to lay straight' despite semantic difficulties).

**dokërr** f 'big bone, bone of arm or leg'

Derived from *\*dok* (for the formation pattern cf. *kokërr*), borrowed from Gk *δοκός* 'rafter, beam'. ◇ Camarda 1864a:85 (to Gk *δόκανα* 'a structure of two joined upright bars'); Meyer 1891:70 (to Turk *dogru* 'direct'); Barić 1919:8 (from *\*dorkr-* composed of *dorë* and *krah*); Çabej 1976a:132 (an expressive form compared with *doçkë* 'little hand' and the like).

**ec** 'to go, to run'

From *\*etës* < PALb *\*aitatja* based on an unattested deverbative nomen actionis *\*aita* < IE *\*oitos* based on *\*ei-* 'to go' similar to Gk (Hom) *οἶτος* 'fate' believed to be connected with *\*ei-*. ◇ Camarda 1864a:95 (to Gk *εἶμι* 'to go'); Meyer 1891:97 (from Lat *\*itiō* replacing *\*itō* 'to go' - but the vocalism remains unexplained); Barić 1919:18 (to aor. *erdha*); ; Jokl *apud* Çabej 1976a:158 (related to *hedh*); Frisk 1972:370-371; Çabej 1976a:157 - 158 (reconstructs *\*itiō* as a source).

**enjë** f 'juniper, yew'

Another variant is *venjë* displaying a phonetically secondary initial *v-*. From PALb *\*aignjā* related to the Indo-European, and in particular Germanic, word for 'oak': ON *eik*, OHG *eih*. ◇ Çabej 1976b:281 (to Lat *acus* 'needle', Lith *aštrūs* 'sharp').

**forbël** f 'peelings, sweepings (of nuts), empty nut-shell'

Other (more archaic) variants are *formël* and *forlë* < *\*formlë*. Its older meaning seems to be 'nut-shell'. The word was borrowed from Lat *formella* 'small form'. ◇ Camarda 1864b:64 (compares *formël* with Gk *φορμός* 'basket'); Meyer 1891:110 (derives *forbël* from *\*vorbël* < Lat *\*orbulus* and *formël* from Ital *forfore* 'scabs'); Çabej 1976a:192 - 193 ("of unclear origin").

**garbë** f 'notch, nick'

Goes back to PALb *\*garbā* etymologically related to OIr *gerbach* 'wrinkled', ON *korpna* 'to get wrinkled', OPrus *garbis* 'mountain' and the like. ◇ Fraenkel 1962:135.

**gargull** adv. 'full'

From PALb *\*garg-ula*, originally a noun related to Lith *gaĩgalas*, *gargõlas* 'thickening, knotted thread, thread'. ◇ Fraenkel 1962:134.

**gath** m 'catkin'

A diminutive in *-th* of an unattested *\*gat* borrowed from Romance *\*gat(t)us* 'cat', cf. Ital *gatto*, Friul *g'at*, Prov *gat* and the like in contrast to *\*cattus* reflected in French *chat*. For the meaning cf. German *Kätzchen* and English *catkin*.

**gëlbazë** f 'liver illness of sheep caused by worms'

Another variant is *këlbazë*. Borrowed from Slav *\*кѣлбаса* 'stuffed gut, sausage', a derivative of *\*кѣлбъ* 'stomach (of animals)'. The irregular change of Slav *\*-s-* > Alb *-z-* is explained by the analogical influence of suffixal forms in *-az(ë)* and *-ëz(ë)*. Rum *gälbează*, *cälbează* has been borrowed from Albanian. ◇ Meyer 1891:222 (to *qelb*); Trubachev 1987:178 - 183.

**gravë** f 'cave, den, lair'

From PALb *\*gravā* etymologically identical with Latv *grava*, *gřava* 'ravine, precipitous valley' further connected with Lith *grĩũti* 'to decline, to collapse', Latv *grūt* id. ◇ Fraenkel 1962:171.

**grumbull** m 'heap, crowd'

Another variant is *grumull*. Continues PALb *\*grumbula* etymologically comparable with Lith *gruĩbulis* 'hump, uneven place' and its cognates connected with *grũblas* 'uneven place, hillock'. ◇ Meyer 1891:132 (from Ital *grumolo* 'cabbage-stump'); Meyer-Lübke 1904:1049 (from Lat *grumus*); Fraenkel 1962:172 - 173.

**gjaj** 'to resemble, to be like; to suit, to become; to seem; to happen'

Dialectal forms *glaj*, *gëljaj* require the reconstruction of PALb *\*ga-lanja* < *\*ga-lab-nja*, a denominative verb based on *\*lab-* etymologically identical with Lith *lābas* 'good', Latv *labs* id. Thus, the original meaning must have been 'to suit, to become'. Note another verbal form *gjas* 'to resemble' also belonging here and continuing *\*ga-latja*. ◇ Camarda 1864a:336 (to Gk γλαύσσω 'to shine', an obvious derivative of γλαυκός 'shining'); Meyer 1891:137 (related to *qas*), 1895:79 (to Gk βάλλω 'to launch, to reach', Skt *galati* '(he) drops, falls down'); Jokl *apud* Çabej 1976a:221 (compares with German *glänzen* 'to shine'); Fraenkel 1962:327; abej 1976a:221 (reconstructs *\*ga-laig-* and links it to Goth *galeikan* 'to please' but this ablaut grade is unknown in *\*leig-* ~ *\*līg-*).

**gjedh m** 'cattle'

From PALb *\*sada* or *\*seda*, a deverbative based on IE *\*sed-* 'to go, to walk'. Semantically, cf. other descriptions of cattle as 'walking', i.e. movable: Gk πρόβατα 'cattle, sheep', Hitt *ijant-* 'ram' and the like. ◇ Pokorny 1959:887; Çabej 1976a:223 (to IE *\*g<sup>h</sup>ōu-* 'cattle' and in particular to Slav *\*govędo*); Benveniste 1969:37 - 45.

**gjemb m** 'thorn'

A Greek-Albanian form *glëmb* preserves the original anlaut *gl-*. Goes back to *\*glamba*, comparable with Slav *\*glqb-okъ* 'deep' < *\*hollowed*, *\*glqbъ* 'trunk, stump, cabbage-stump'. ◇ Meyer 1891:140 (to Lith *gėmbė* 'nail used to hang clothes' - impossible in view of the initial *gl-*); Jokl 1911:26 - 28 (to Lith *gėliū*, *gėlti* 'to stick'); Trubachev 1979:141 - 143.

**gjezdis** 'to go for a walk, to roam'.

An early borrowing from Slav *\*jězditi* 'to ride' with the initial *j-* substituted by Alb *gj-*, cf. South Slavic continuants: Bulg *zplz*, SCr *jezditi*.

**gjije f** 'stable, house'

A singularized plural of a form attested in Geg as *gjê* 'stable, pen'. Goes back to *\*saina* identical with the Baltic word for 'wall': Lith *siena*, Latv *siena*. ◇ Fraenkel 1962:782 - 783; Çabej 1976a:228 (important lexical material but no etymology).

**gjurmë f** 'trace'

From PALb *\*surma*, a zero-grade variant of IE *\*sor-mo-* reflected in Skt *sárma-* 'flow', Gk ὄρμη 'assault, attack', further connected with IE *\*ser-* 'to flow'. ◇ Meyer 1884:59 (borrowed from Romance via NGk γούρμα id.), 1891:142 (uncertain link to Ital *orma*, Rum *urma*); Barić 1919:103 (to Lat *serpō*).

**References**

Barić, Henrik

1919 *Albano-rumänische. Studien I.* Sarajevo: Institut für Balkanforschung.

- 1954 *Lingvističke studije*. Sarajevo: Naučno Društvo NR Bosne i Hercegovine.
- 1955 *Hymje në historinë e gjuhës shqipe*. Prishtinë: Mustafa Bakija, 1955.
- Benveniste, Emile  
1969 *Le vocabulaire des institutions indo-européennes* I. Paris: Klincksieck
- Çabej, Eqrem  
1976a,b *Studime gjuhësore* I-II. Prishtinë: Rilindja.
- Camarda, D.  
1864a, b *Saggio di grammatologia comparata sulla lingua albanese*. Livorno: Successore di Egisto Vignozzi.
- Fraenkel, Ernst  
1962-1965 *Litauisches etymologisches Wörterbuch*. I - II. Heidelberg - Göttingen: Carl Winter and Vandenhoeck & Ruprecht.
- Frisk, Hjalmar  
1960-1972 *Griechisches etymologisches Wörterbuch* I - II. Heidelberg: Carl Winter.
- Huld, Martin  
1984 *Basic Albanian Etymologies*. Columbus: Slavica.
- Jokl, Norbert  
1911 *Studien zur albanesischen Etymologie und Wortbildung*. Wien: A. Hölder.
- 1923 *Linguistisch-kulturhistorische Untersuchungen aus dem Bereiche des Albanischen*. Berlin - Leipzig: Walter de Gruyter.
- Mayrhofer, Manfred  
1953-1980 *Kurzgefaßtes etymologisches Wörterbuch des Altindischen*. I - IV. Heidelberg: Carl Winter.
- Meyer, Gustav  
1884 *Albanesische Studien II. Die albanesischen Zahlwörter. Sitzungsberichte der philosophisch-historischen Classe der Kaiserlichen Akademie der Wissenschaften CVII* (1884), 259 - 338.
- 1891 *Etymologisches Wörterbuch der albanesischen Sprache*. Straßburg: Karl J. Trübner.
- 1895 *Albanesische Studien V. Beiträge zur Kenntniß der in Griechenland gesprochenen albanesischen Mundarten. Sitzungsberichte der*

*philosophisch-historischen Classe der Kaiserlichen Akademie der Wissenschaften* 134/7.

Meyer-Lübke, Wilhelm

1904 Albanisches. In: *Grundriß der romanischen Philologie*. I - II. Straßburg: Karl J. Trübner, 1888 - 1902.

Miklosich, Franz

1871 *Albanische Forschungen. II. Die romanischen Elemente im Albanischen*. Wien [= *Denkschriften der Kaiserlichen Akademie der Wissenschaften* XX].

Orel, Vladimir

1996 *Albanian Etymological Dictionary*. Jerusalem[ms.].

Pokorny, Julius

1959 *Indogermanisches etymologisches Wörterbuch*. I. Bern: Francke.

Tagliavini, Carlo

1937 *L'albanese di Dalmazia*. Firenze [= *Biblioteca dell'Archivum Romanicum* XXII, ser. 2].

Trubachev, Oleg N. (ed.)

1975-1987 *Étimologičeskii slovar' slavianskikh iazykov* II - XIII. Moscow: Nauka.

Vasmer, Max

1921 *Studien zur albanesischen Wortforschung*. Dorpat: Universität Verlag.

1970 *Étimologičeskii slovar' russkogo iazyka* II. Moscow: Progress.



## Macro Families: Can a Mistake Be Detected?

Ilia Peiros

*The University of Melbourne*

One of the linguistic courses I did at University about 25 years ago was a course of Nostratics taught by Aaron Dolgopolsky. At that time I found myself rejecting Nostratics completely, due to the paucity of my knowledge of the language families discussed: "Why should I accept this strange idea? Comparative Slavonic is much more convincing and I know the languages involved". I also could not understand why Dolgopolsky, an outstanding linguist and respected scholar spent his time on such an odd theory rather than doing something more convincing. A few years later, I joined the Nostratic Seminar, not from a love of Nostratics but because I liked the people involved: Dybo, Dolgopolsky, Starostin, Khelimsky and many others. Through this opportunity for observation and discussion, I reached some understanding of how the Nostraticists work, mainly what is the theoretical background of long-range comparisons. In this article I discuss only procedures for assessing claims that languages are related<sup>1</sup>, and apply them to the study of East and Southeast Asian languages.

A hierarchy of least six different levels can be distinguished in comparative linguistics: Dialect, Language, Young Family, Developed Family, Old Family and Macro Family (cf. Jakhontov 1980). The distinction between these linguistic levels is not absolute, and in many cases it is hard to tell which level a particular genetic unit belongs to. Still, the distinctions are useful, and are worth including in further discussion. The units can not be defined in a formal way, but have some specific features:

- Dialect.

People who speak the same dialect understand each other without any restrictions. Differences in their speech varieties are mostly explained by the existence of sociolinguistic factors known to the speakers. Normally there are no doubts that a dialect represents a single genetic entity.

- Language.

It is important to distinguish the two notions, 'language' and 'sociolanguage'. Although two speakers of the same language may use different dialects with sometimes quite noticeable differences they will always understand each other if they discuss common topics. The criterion of mutual intelligibility is thus essential for the notion 'language'. However, two speech varieties are included in one sociolanguage if the speakers believe that they speak the same (socio)language, regardless of their actual ability to understand each other. Languages and sociolanguages form different combinations:

---

<sup>1</sup> The procedures are based on discussions which were conducted at the Nostratic Seminar during the 70s and 80s. Their formulation, however, took me another ten years, and I alone carry the responsibility for their current presentation.

One language – one sociolanguage

All speakers of Hungarian know that they use the same (socio)language. Slight differences in dialects do not prevent mutual intelligibility, which means that they belong to the same language.

One language – two or more sociolanguages.

This situation is represented, for example, by Serbian and Croatian: mutual intelligibility (one language), but people know that they speak different (socio)languages.

Several languages – one sociolanguage.

The Chinese 'dialects' present the best example of this type. It is well known that the differences between some of them are not less than between Slavonic languages and in many cases mutual intelligibility is not possible. On the other hand, speakers of Chinese 'dialects' know that they speak the same (socio)language and this fact often affects their behaviour.

In comparative linguistics the main focus is on languages, not sociolanguages.

People who speak different dialects of the same language can identify differences between their speech varieties and often like to discuss them. It is important that they talk about differences rather than similarities the existence of which do not surprise them at all. Genetic relationship of two dialects which belong to one language is always self evident for both speakers and linguists.

• Young Family.

Speakers of two languages which belong to a Young Family, for example Russian and Ukrainian, are able to communicate, but usually only with difficulty. The speakers can maintain conversation but it will often be interrupted by repetitions and changes. In many cases it is hard to say if we are dealing with two dialects of the same language or with two languages included in one Young Family. Genetic identity of the members of one Young Family is always clear for speakers and linguists. Speakers usually pay more attention to differences between their speech varieties rather than to similarities between them.

• Developed Families.

The Slavonic or Germanic languages provide examples of families which belong to this level. Normally, speakers of two languages which belong to the same Developed Family do not understand each other. However, they can find many similarities between any two members of such a family, say Russian and Czech, English and Danish, etc. Comparing such languages, a speaker would talk not about differences, which are taken for granted, but about similarities between them. Speakers' explanations of such similarities may vary, but quite often they are based on an assumption of common origin of the languages.

For comparative linguists, genetic relationship of members of Developed Families is usually self evident and does not cause any problems apart from the questions of classification. Data from Developed Families is quite transparent, and normally it is not too difficult to connect a reconstructed proto-form with its reflexes in recorded languages.

Most comparative linguists conduct their research at the level of Young or Developed Families. Quite often such linguists have native or near native command of several languages of the family, which allows them to operate with high quality first-hand data.

A reconstruction of a Developed Family's proto-language is based on comparison between spoken or written languages:

Language A      =>  
 Language B      => Proto-language DF  
 Language C      =>

Sometimes, however, an intermediate reconstruction is needed if a Developed Family includes a Young Family as one of its branches.

• Old Families.

A good example of an old language family is Indo-European, which includes, among many other languages, French and Hindi. French speakers who study Hindi normally cannot identify forms of common IE origin, retained in both languages. Various changes which took place in the history of these languages have resulted in originally similar forms becoming absolutely different. For this reason, speakers cannot create any reasonable hypothesis about the genetic affiliation of modern languages which belong to different branches of an Old Family: their intuition does not work at this level. For comparativists, on the contrary, it is quite normal to deal with Old Families and their validity as families is widely accepted. It is important, however, that the existence of the IE family has been discovered not through comparison of modern languages like French and Hindi, but through knowledge of ancient languages like Latin, Greek and Sanskrit (which are much closer to each other and, I think, actually belong to the same Developed Family, as the similarities between them are easy to detect).

The Proto-IE language was originally reconstructed mainly through direct comparison of ancient languages, which meant that the work occurred at the level of Developed and not Old Families. Evidence from Proto-Germanic, Proto-Slavonic, Proto-Celtic, and other studies was subsequently incorporated into Proto-IE investigation, which nowadays is reoriented toward comparisons of reconstructed Proto-languages rather than simply archaic languages:

Language A	=>	{	Proto-language DF(A)	=>	{
Language B	=>				
Language C	=>				
Language D	=>	{	Proto-language DF(K)	=>	{
Language E	=>				
Proto-language OF(AR)					
Language G	=>				
Language H	=>	{	Proto-language DF(R)	=>	{
Language I	=>				
Language J	=>				

Theoretically, it is clear that intermediate proto-languages rather than recorded ones should be used in the reconstruction of Old Families. Each intermediate proto-language is associated with a Young or Developed family, which includes transparently related languages. If ancient and archaic languages are not known, success in the study of an Old Family depends on the existence of such intermediate reconstructions.

To do research at the level of an Old Family a linguist needs to be familiar, not only with the main languages of the family, but also with problems in the reconstructions of all proto-languages constituting the family. The amount of information required is many times more than in a study of a Developed Family, and this automatically reduces the number of specialists in Old Families. I do not know the exact number, but I estimate it to be roughly ten to twenty times less than the number of people studying Young or Developed families.

At the level of Old Families we face for the first time a conflict between the judgments of speakers and linguists: the latter recognise the genetic relationship, while the former do not.

• Macro families.

Comparing even very archaic languages which belong to different branches of a Macro Family, we will normally find only a limited number of similarities, which often are not fully convincing. That is why the justification of macro-families such as Illich-Svitych's Nostratics, and Starostin's Sino-Caucasian, is based not on comparisons of recorded languages, but on reconstructions of proto-languages constituting them, which in their turn can be based on reconstructions of younger proto-languages. A Macro Family reconstruction is therefore based on several levels of intermediate reconstruction conducted independently for each of its daughter proto-languages. Three or four levels of reconstruction are quite normal at this level of complexity:

I

Language A	=>	{	Proto-language DF(AC)
Language B	=>		
Language C	=>		
Language D	=>	{	Proto-language DF(DG)
Language E	=>		
Language G	=>		
Language H	=>	{	Proto-language DF(HJ)
Language I	=>		
Language J	=>		
Language K	=>	{	Proto-language DF(KM)
Language L	=>		
Language M	=>		
Language N	=>	{	Proto-language DF(NP)
Language O	=>		
Language P	=>		
Language Q	=>	{	Proto-language DF(QS)
Language R	=>		
Language S	=>		

Language T	=>	{	Proto-language DF(TV)
Language U	=>		
Language V	=>		

Language W	=>	{	Proto-language DF(WY)
Language X	=>		
Language Y	=>		

## II

Proto-language DF(AC)	=>	{	Proto-language OF(AH)
Proto-language DF(DG)	=>		
Proto-language DF(HJ)	=>		

Proto-language DF(KM)	=>	{	Proto-language OF(KQ)
Proto-language DF(NP)	=>		
Proto-language DF(QS)	=>		

Proto-language DF(TV)	=>	{	Proto-language OF(TW)
Proto-language DF(WY)	=>		

III	Proto-language OF(AH)	=>	{	Proto-language MF (AW)
	Proto-language OF(KQ)	=>		
	Proto-language OF(TW)	=>		

The Study of Macro Families is characterised by:

(i) the absence of archaic languages which could be directly compared to justify the relationship. Reliable similarities are found between reconstructed proto-languages, rather than between any recorded ones.

(ii) lack of transparency in reconstructions: normally one cannot tell whether recorded forms can be traced back to the proposed Macro Family reconstruction. To check this one needs to know the histories of all the daughter-families.

(iii) usually daughter-families are investigated by separate branches of comparative linguistics, with no significant tradition of cross reference.

These three features make any research at the Macro Family level extremely difficult. Because of the huge amount of data required, only people of exceptional ability can work at this level of complexity. I believe that altogether less than 50 scholars in the linguistic world can successfully study Macro Families. This gives them a brilliant opportunity to talk about their hypotheses and problems in an exclusive club ignoring the needs of other linguists. Should we simply wait for revelations issued from this club, or can we investigate their claims ourselves?

There are four major classes of objections to hypotheses about Macro Families:

(i) One can reject the whole idea of long range comparisons as the product of pure imagination and thus beyond true scholarship. Different arguments have been produced to support such an approach: "the comparative method cannot be applied to such remote periods of time", "the languages must have been different in the remote past and thus they would have followed other rules of

development", and many others of this type.<sup>2</sup> I think, however, that the underlying sense of all such claims can be formulated as follows: "I can not accept Nostratics or the Sino-Caucasian hypothesis as they contradict my intuition and they are too complex for me to evaluate. I do not have time to struggle with the evidence presented, and instead I will produce some general suggestions to support my feelings". Such claims are not, however, based on any linguistic evidence, and therefore we cannot seriously discuss them.

Leaving aside such general rejections two approaches are related to studies of particular Macro Families.

(ii) One can reject a Macrofamily hypothesis on the grounds that it does not fit with the broader prehistoric picture: "Nostratics is wrong, as it is not supported by extra-linguistic evidence", "I cannot fit the Sino-Caucasian theory into my understanding of Asian prehistory", and so on. It is clear, however, that such considerations can be used in the historical interpretation of linguistic hypotheses, but are not applicable to the discussion of these hypotheses within linguistics.

(iii) More serious objections come from the following direction. Talking about a Macro Family, a specialist in history of a particular Old Family conducts a thorough investigation of relevant reconstructions used in justification of the Macro Family. Taking, say Indo-European data in Nostratics, the linguist finds several wrong or unconvincing IE proto-forms included in Nostratic etymologies. This fact leads to the conclusion the whole Nostratic hypothesis "remains as yet a house of cards" (Vine 1991:31). Logically speaking, this is one of two possible ways to reject a Macro Family claim. To use this option properly, however, one needs to demonstrate that all or most Indo-European comparisons used in Nostratics are wrong.<sup>3</sup> If there are only a small number of incorrect Indo-European forms, they can be removed from Nostratic etymologies without destroying the whole hypothesis.

(iv) Another way to reject a Macro Family hypothesis is based on the analysis of methods used in its justification. It could be argued that the study of Macro Families belongs to comparative linguistics, and the same methods and procedures should be used in investigation of any language families: Young, Old or Macro. If we could demonstrate that the methods of comparative linguistics were violated, that would immediately take the corresponding hypothesis out of the discussion: all statements of comparativists have their meanings only within comparative linguistics.<sup>4</sup>

It seems to me, that the method of comparative linguistics provides us with reliable tools for the formal evaluation of any claims about genetic

---

<sup>2</sup> A example of such reasoning is represented by the following remark of an anonymous internal referee of Peiros, to appear: "Peiros believes that the step from Proto-Indo-European to "Proto-Nostratic" is just as straightforward as, eg. from Proto-Indo-Iranian to PIE. Does not seem to realise that after a certain period in time (ca. 10,000 years B.P. at a very generous estimate) genetic relationship is indistinguishable from borrowing or pure chance."

<sup>3</sup> It seems that this can be done within Japanese / Austro-Tai hypotheses for its Japanese and Miao-Yao components (see below).

<sup>4</sup> Note, however, that comparative linguistics often shares terminology with other theories, which use totally different methods.

relationship of languages and thus for evaluation of the validity of claims about Macro Families. The comparative method, as it can be described now, is based primarily on the study of rather transparently related languages which usually belong to Young or Developed families. Old families are also studied at a level of transparency, but reconstructed instead of recorded forms are used. If we want to study even more ancient genetic units - Macro Families - we also need to operate with transparent relationship, which means that only a study based on reconstructed potential siblings is acceptable. In other words, to deal with a Macro Family we need to base our arguments on constituent proto-languages which reveal simple and transparent relationships. The collected data should be able to convince linguists who are not specialists in this particular theory, and who play here the same role as untrained speakers at the level of Young and Developed families.

Taking at random two languages A and B we could reach one of three possible conclusions as to their genetic relatedness:

(i) they are genetically related and share the same ancestor. Their relationship may be transparent as on the level of Young or Developed Families, or it could be more obscure, as between members of major branches of Old and Macro families;

(ii) language A is a direct or remote ancestor of B;

(iii) it is not known if A and B are genetically related. In such cases linguists would usually say that the languages are not genetically related, despite the fact that within comparative linguistics it is impossible to demonstrate the absence of genetic relationship. The only theoretically correct claim which can be made is that there is no available evidence that the two languages are related.

Let us limit our discussion to the first possibility: genetically related languages. It is generally accepted that languages are genetically related if they can be traced back to the same common ancestor. This means that strictly speaking if we want to demonstrate that languages A and B are related, we must present their ancestor, language C. With few exceptions C would be a proto-language, whose system is reconstructed through the comparative method based on a comparison of its daughter-languages. This leads us to a vicious circle: to prove that the languages are related we need a reconstructed proto-language, but to reconstruct it we need to know which languages are related. To overcome this contradiction we can use a working definition of genetic relatedness which does not include the notion of proto-language.

Related languages usually contain certain similarities which are traces of their common origin. Such similarities can be functional and / or material. In the case of pure functional similarities, certain parts of linguistic systems are organised similarly. For example, two languages might distinguish identical sets of noun classes, although the grammatical morphemes used to mark the classes could be quite different. Systemic features and their particular combinations do not appear at random, so it is not impossible that these similarities indicate genetic relationship. It is, also highly probable, however, that they are results of areal influences, typological universals and other non genetic factors. For this reason, functional similarities should never be used as the sole piece of evidence for genetic relationship and, in fact, they are not used as such in any well attested case.

The main body of evidence comes from material similarities. These include similarities between morphemes of the languages sometimes together

with similar irregularities found in the languages, such as between English and German irregular verbs: *drink, drank, drunk* vs. *trinken, trank, getrunken*. It is very hard to believe that such irregularities can be borrowed or result from independent development, so they are convincing indicators of possible genetic relationship. Unfortunately, however, they are not found often enough and genetic claims are primarily based on morphemic similarities and conclusions drawn from them. A list of similar morphemes found in the languages under investigation is absolutely crucial, and without it no genetic claims can be substantiated within comparative linguistics.

Before taking the next step in our discussion we need to clarify several notions. If the history and relationships between the languages are known:

- morphemes are called genetically related if they all result from the direct and uninterrupted development of the same morpheme of the proto-language. This morpheme is called their *proto-morpheme (proto-form)*.
- morphemes which can be traced back to the same proto-morpheme are called *cognates*.
- a set of cognates developed from a single proto-form is called an *etymology*. An etymology thus includes only genetically related forms found in the languages under investigation.
- morpheme *a* in language *A* is a *reflex* of the proto-morpheme *\*f*, if *a* is a result of direct development of *\*f* in the history of *A*; phoneme *a* in language *A* is a *reflex* of the proto-phoneme *\*f*, if *a* is a result of direct development of *\*f* in the history of *A*.

If the history and interrelationship between the languages is not yet known:

- similar morphemes in those languages are called *resemblances*. There are various reasons why the morphemes may be similar: they could be cognates, borrowings, or even chance similarities.
- A set of resemblances found in the languages is called a *comparison*. No substantial claims can be made about the origins of a comparison. An etymology is a particular type of comparison, one which includes only genetically related morphemes.

Now using the notions of an etymology (= a set of genetically related cognates) and a comparison (= a set of resemblances which are not necessarily genetically related) we can suggest a working definition of genetic relationship. Languages are genetically related if:<sup>5</sup>

(i) there is a sufficient number of comparisons consisting of resemblances found in these languages;

(ii) it can be demonstrated that these comparisons are etymologies in the strict sense and not borrowings or chance similarities. As the only accepted way to demonstrate the genetic nature of a comparison is to show that its resemblances are connected by systemic phonological correspondences (reflections of certain features of the proto-language phonological system), a list of systematic phonological correspondences is another necessary element for proof of genetic

---

<sup>5</sup> This definition does not specifically require identity of grammatical morphemes. It is based on my experience in comparative study of South-East Asian languages, which usually do not have developed grammatical systems, but still obviously form clear-cut genetic units.



relationship. In many cases the systematic correspondences also help us to identify loans.

Fulfilling these two conditions for the demonstration of genetic relationship provides us with information sufficient for phonological and lexical reconstruction of the proto-language. For the families with old morphology we should also be able to reconstruct common grammatical morphemes on the basis of comparisons between daughter language grammatical morphemes. According to the given definition, however, reconstructions are not required for proof of genetic relationship.

This definition is designed to work in the cases of transparent genetic relationships, like those represented in Young and Developed families, which are supported by the intuition of both speakers and linguists. It does not work directly for Old or Macro families, where instead of modern languages, one must work with their archaic ancestors, recorded or reconstructed. In research at this level it is still important to operate with rather transparently related languages and to apply the same two conditions, treating these (proto-)languages in the same way as modern ones.

In this paper I want to discuss the theoretical validity of evidence presented to support the following genetic claims: Sino-Caucasian, Japanese / Austro-Tai, Sino-Austronesian and Miao-Yao-Austroasiatic. The Sino-Caucasian theory (Starostin 1982, 1984) claims that three language families, Northern Caucasian, Eniseian and Sino-Tibetan are genetically related. The Japanese / Austro-Tai (JAT) theory claims the Austronesian languages are related to Kadai, Miao-Yao and Japanese (Benedict 1990). According to the Sino-Austronesian theory, Chinese and the Austronesian languages are genetically related (Sagart 1993; 1994), which contradicts both the SC and the JAT hypotheses. The Miao-Yao / Austroasiatic claim (Jakhontov 1981, Peiros to appear) connects Miao-Yao with the Austroasiatic rather than with Kadai and Austronesian families.<sup>6</sup>

Based on the requirements outlined in the working definition above, we can subdivide these genetic claims into three groups:

1. well supported claims: a sufficient number of comparisons, connected by systematic phonological correspondences are presented. Whether the genetic relationship is established depends on the quality of data presented, but the formal requirements of the definition are fulfilled. Starostin's SC theory belongs here: a list of comparisons is given, and major phonological correspondences are established. Strictly speaking, only claims of this type can be fully discussed and formally evaluated within comparative methodology.

2. plausible claims: these are based on a certain number of comparisons, but no systematic correspondences are established. Such claims are often just indications that further research is needed to 'upgrade' their level with more similarities and a set of phonological correspondences. I believe that the Kadai - Austronesian and Miao-Yao - Austroasiatic hypotheses belong here.

3. claims not supported by convincing evidence: the comparisons given do not necessarily indicate genetic relationship and no set of phonological

---

<sup>6</sup> I argue (to appear) that Kadai-Austronesian and Miao-Yao-Austroasiatic are two main branches of the Austric macrofamily.

correspondences is provided. I think that the treatment of Miao-Yao and Japanese in JAT and the whole Sino-Austronesian hypothesis belong here.

In evaluating genetic claims, we need to consider the following issues:

1. initial data: which languages in which form are compared;
2. how comparisons are identified;
3. quality of phonological correspondences, if any.

#### Initial data.

Two types of data can be used in justifying a genetic claim: reconstructed proto-forms and morphemes taken from recorded languages. At least five types of proto-forms can be found in the literature: real reconstructions, areal reconstructions, reflections, pre-reconstructions and ghost-reconstructions (This rather vague terminology is mine). The most reliable are *real reconstructions*. They are obtained through the strict universal procedure of comparative linguistics: (i) their identification is based on the system of phonological correspondences and plausible semantic relationships, (ii) their reflexes are found in all or major languages of the family and (iii) they can be definitely attributed to the proto-language level.

Sometimes a reconstruction is based on forms found in several languages and is confirmed by proper phonological correspondences, but it cannot be demonstrated that the form should be attributed to the proto-language level, rather than to a later period of the family's history. In such cases we are dealing with an *areal reconstruction* which could belong either to the proto-language of the whole family, to one of its daughter proto-languages, or represents unidentified areal influences on some languages of the family. The status of areal proto-forms is similar to that of proto-forms reconstructed for different branches of the family: in both cases we do not know at exactly what level of relationship they represent. Their usage undermines the validity of a genetic claim.

Many proto-forms used in the justification of JAT are areal reconstructions. Among them are all PAN forms based solely on Formosan data. The AN languages of Taiwan reveal similarities with Kadai and Japanese, which are not found in AN languages elsewhere. There are two possible explanation for this. One can assume that the Formosan languages have retained a significant number of PAN forms lost in other languages of the family (this is the position of Benedict), or it can be suggested that these languages preserve traces of contact with Kadai and/ or Japanese which took place after the disintegration of PAN (the geographical position of Taiwan makes this suggestion rather convincing). As we do not have enough data to choose between the two options, it would be better not to use these areal reconstructions in justifying this genetic claim.

If a morpheme is recorded in a language with a known history, but cognates are not found in other related languages, a linguist who believes that this morpheme is not a borrowing can assume that its ancestor form was also represented in the proto-language, and a corresponding proto-form can be reconstructed. Such "*reflections*" have less convincing power than real reconstructions, as there are no general reasons why they should be attributed to the proto-language level rather than to the level of one of its daughter (proto-) languages.

Reflections are often used in JAT: PAN *\*ʔumuq* 'pus' based only on Paiwan *umuq* (Benedict 1990: 232) is an example (if the phonological relationship between the Paiwan and Proto AN forms can be accepted). Obviously it is very hard to demonstrate that a reflection really belongs to the proto-language and is not, say, a later borrowing. Reflections, however, can be used in comparisons, providing they do not form the major body of evidence.

Two other types of proto-forms found in the literature, pre-reconstructions and ghost-reconstructions, do not, strictly speaking, belong to comparative linguistics. *Pre-reconstructions* are not based on a proper set of phonological correspondences but only on the intuition of the linguist who introduced them. Working with language families for which the comparative phonology is not well known a linguist may bring together fragments of historical information to gain an idea of how a proto-form might look. To transfer such pre-reconstructions into real reconstructions a detailed comparative phonology of the language family under consideration is needed. Without it, any genetic claim based on similarities between pre-reconstructions remains only a hypothesis.

The degree to which pre-reconstructions are convincing depends on such factors as:

(i) development of comparative phonology. For example, the main features of the PAN phonological system are known, but additional study is needed to work out detailed histories of its daughter families and their constituent languages. Any PAN form which is based on new data from languages with relatively obscure phonology, like Tsouic or Atayal, remains a pre-reconstruction (although perhaps a plausible one).

(ii) similarity between (proto-)languages. Forms of different Kadai branches are often rather similar, which in some cases makes pre-reconstructions quite convincing.<sup>7</sup> By contrast, relationships between Miao-Yao languages are much more obscure and the Proto-Miao-Yao pre-reconstructions used by Benedict are consequently less reliable.<sup>8</sup>

A *ghost-reconstruction* is the most treacherous type of proto-form found in the literature. It is usually based on a single morpheme, sometimes marginally represented in a language or simply on a mistake due to poor knowledge of the language's history. Four different Proto Miao-Yao ghost-reconstructions, each supported by a single form from only one Yao dialect can be found for example in the JAT comparison HOLD/ BITE/ CHEW (Benedict 1990: 209-211): *\*khamgam<sup>B</sup>* (< Haininh Mun Yao *khamgam<sup>B</sup>* 'jaws'), *\*ngam<sup>C</sup>* (< Haininh Mun Yao *gam<sup>C</sup>* 'press with the hand, crutch'), *\*gom<sup>C</sup>* (< Chianrai Mien Yao *kom<sup>L</sup>* 'to fetter, shackle') and *\*ngom<sup>A</sup>* (< Haininh Mun Yao *geom<sup>A</sup>* 'hold in mouth'). No conclusions can be made on the basis of such ghosts.

Distinguishing these five types of proto-forms allows us to describe genetic claims as comparative (based primarily on true reconstructions) or

<sup>7</sup>In fact, the relations between Kadai languages are more complicated than they seem at first, and to work out a Proto-Kadai phonological reconstruction is a challenge (Peiros, to appear).

<sup>8</sup>My Miao-Yao reconstructions are based on a set of systematic correspondences between Proto Miao and Proto-Yao and forms represented in both branches of this family (Peiros, to appear).

heuristic (based on pre-reconstructions). Only comparative claims can formally justify a genetic relationship. Starostin's SC theory exemplifies the first type of claim, the other hypotheses mentioned above are all heuristic rather than comparative.

It is absolutely clear that the reconstructions used in a comparison should be self-reliant, which means that they should be obtained independently from each other, and not 'tuned' for better similarity. If a new version of a reconstruction is used only to support a genetic claim, we should be quite suspicious: often it means that the proto-forms have been 'tuned'. The most secure cases are when proto-forms are taken from already existing sources, comparative dictionaries or reconstructions made beforehand,<sup>9</sup> rather than suggested in the publication which makes the genetic claim. It seems to me that *a priori* a genetic claim based on previously known proto-forms is much more convincing than a claim based on proto-forms created especially for the purposes of its justification. That is why I still believe that the Benedict's original AT article (1942) is more convincing than his whole AT book (1975). Real reconstructions, by their nature, can not be 'tuned'; this is the 'privilege' of pre-reconstructions and ghosts.

In contrast to proto-forms, most recorded morphemes and words are real and reliable. In many languages, like Chinese or Japanese, due to various losses and mergers a modern form could be traced back to many different ancient forms. Only thorough investigation of the language's history can reveal its real ancestor. Such investigation is usually based on detailed study of historical phonology and lexicology. That is why I personally always have strong suspicions when external comparisons are based on a new version of the historical phonology of a language. Much more reliable are forms taken from historical dictionaries or phonological studies, rather than those adjusted for external comparisons. Only in the first case can one be sure that the forms are reconstructed properly. That is why it is appropriate to treat the Japanese forms in JAT with suspicion: the sources of Old Japanese forms in JAT often remain to obscure and in many cases are not supported by the history of Japanese (Vovin 1994: 373-376). In SC theory, however, all Archaic Chinese forms are taken from Starostin's Archaic Chinese reconstruction, completed much earlier than the SC studies began.<sup>10</sup>

In order to fully illustrate the effect that quality of initial data has on whether a genetic claim is convincing or not, it is worth comparing in detail Starostin's SC and Benedict's JAT theories.

The SC theory is based on the comparison of Proto North-Caucasian, Proto Eniseian and Proto Sino-Tibetan reconstructions which have been made absolutely independently from each other. In reconstructing Proto NC and Proto EN, the precise method of comparative linguistics was used including step by step movement from recorded languages towards their common ancestor. Several intermediate reconstructions, such as Proto-Lezghinian and Proto-

<sup>9</sup> The SC theory was originally discussed in the article which also included the Proto-Eniseian reconstruction. This reconstruction is, however, self-sufficient and is not based on data from other language families.

<sup>10</sup> The reconstruction has been published in 1989 (Starostin 1989), but it was completed much earlier.

Dagestanian, were created before dealing with Proto NC. Each of these reconstructions is based on a set of systematic phonological correspondences and a representative list of etymologies. An NC comparative dictionary is now published (Nikolaev and Starostin 1995). Proto-EN etymologies are given in the first part of Starostin 1982, with data demonstrating that the proto-forms are based on phonological correspondences and are not 'tuned'.

The situation with ST reconstructions is more complicated. When Starostin published his SC comparisons, the following ST data was available to him: Benedict's Tibeto-Burman reconstructions, Starostin's own Archaic Chinese reconstruction and our comparative ST (Peiros and Starostin 1996) dictionary which was unpublished in that time. To make his SC results more convincing, Starostin chose to use Benedict's proto-forms together with his own Archaic Chinese forms, rather than quote from the unpublished dictionary. Data from the dictionary was used indirectly: only comparisons accepted in it were included in the SC etymologies. This approach, however, had a weakening effect on the whole theory, as Benedict's proto-forms are not true reconstructions supported by a complete set of systematic phonological correspondences but only pre-reconstructions reflecting Benedict's historical guesses.

The proto-forms used for justification of the JAT hypothesis are of quite different nature. To illustrate this, we can draw examples from Benedict's treatment of each of the four language families involved: Kadai, Miao-Yao, Austronesian, and Japanese.

The Kadai family includes 6 to 8 branches, with proto-languages reconstructable for at least three of them (Zhuang-Tai, Kam-Sui and Li). A comparison of these proto-languages with genetically isolated Ong Be and Likkja leads to Proto-Kadai, the reconstruction of which is not yet published (Peiros, to appear). 'Tuned' proto-forms from three intermediate reconstructions – Zhuang-Tai (or Tai) by Li Fangkuei (1977), Kam-Sui (Thurgood 1988) and Proto-Li (Matisoff 1988) – are used in Benedict's 1990 book. A good example is Benedict's Proto-Kadai form 'sugarcane'  $*[t]o[b]oi > *[t]o[w]oy > *C_{\text{tooy}}^B$  based on Zhuang-Tai  $*_{\text{tooy}}^B$  'sugarcane', Kam-Sui  $*_{\text{tooy}}^B$  'sugarcane' and Southern Li (which dialect?)  $oi^C$  'maize' (1990, 232). It is based simply on obvious similarity between Zhuang-Tai and Kam-Sui forms and the need to connect them with Proto Austronesian  $*təbus$  'sugarcane'. This, and all other Kadai proto-forms discussed by Benedict, remain to be pre-reconstructions. Most Kadai languages are quite similar to each other (they perhaps form a Developed Family), and usually it is not too difficult to identify comparisons. However, intensive internal contacts and impact from Chinese, Vietnamese, Khmer and other Southeast Asian languages necessitate a detailed knowledge of Kadai comparative phonology for proper genetic interpretation of comparisons found.

The Miao-Yao family with its two main branches (Miao and Yao) presents another type of problem. Due to the occurrence of significant phonetic changes, identification of comparisons even in closely related Miao languages can be quite challenging, especially if forms are not known from more archaic dialects (languages). Phonological correspondences connect the main Miao dialects (Wang 1985), but a detailed Miao comparative dictionary does not exist. Proto-Yao is known mainly thanks to Purnell's reconstruction (1970), which requires some revision (Peiros, to appear), with extensive reliable data available for only one dialect (Lombard 1968). These limitations mean that Benedict's MY

proto-forms remain pre-reconstructions, less convincing than those suggested for Proto-Kadai. Their reliability is also undermined by possible Chinese borrowings which are hard to detect without proper phonological information.

The Austronesian family includes many hundreds of languages, grouped into various branches and sub-branches, each with its own proto-language. The phonological history of Proto-AN and its main descendants is known much better than that of Kadai or Miao-Yao, so we have quite reliable reconstructions of many PAN morphemes. However, the classification of the family remains quite uncertain. Many linguists accept different versions of Blust's provisional AN classification, which unfortunately is not a purely genetic one.<sup>11</sup> This uncertainty makes it very difficult to demonstrate that a reconstructed morpheme belongs to the proto-language level, rather than to some more recent one. There is no general agreement about how to solve the problem, but all Austronesianists agree that this is not a simple and straightforward task and that each etymology should be thoroughly investigated before it can be called Proto-AN (see, for example Mahdi's (1994) painstaking efforts with a few possible AN etymologies). There is however an extensive collection of AN etymologies published mainly by Dempwolff (1934-38) and Blust (1980; 1983-4, 1986; 1988; 1989), but more than half of all AN etymologies used in Benedict's comparisons are not found in these major sources. Instead he operates with his own pre-reconstructions based largely on the data Formosan languages. This a situation, plus widespread 'tuning' of proto-forms, makes the whole AN part of the JAT theory a collection of data which should be treated with great caution. A sample of AN pre-reconstruction can be \*[C,s]ama 'green' based on a single form: Sediq sama 'green'. Sediq is an Atayalic language of Taiwan whose relationship with Proto-AN remains very obscure. An example of a 'tuned' proto-form is Benedict's \*[q,ʔ]u(n)[z]ay instead of Dempwolff's \*ʔuɖay 'worm' (Benedict 1990: 263). The 'tuning' is needed to justify a comparison with Japanese uzi.

The history of Japanese can be understood only with the help of Old Japanese and Common Japanese-Ryukyuan reconstructions as most of the modern forms can be traced back to several different ancient forms. Intensive studies are undertaken in this area (see, for example, Martin 1987), but Benedict does not follow any particular reconstruction and uses modern Japanese forms or his own Old Japanese pre-reconstructions, which quite often are misleading (Vovin 1994).

#### Identification of comparisons

Given good quality initial data, the next step in checking a genetic claim is an analysis of comparisons included in it, and especially the evidence that these comparisons are real etymologies. The only way to demonstrate the genetic nature of comparisons is to analyse them with the help of systematic phonological correspondences between the languages under investigation and to

---

<sup>11</sup>Blust (1980: 11-12), for example, defines the Western-Malayo-Polynesian group not as a genetic unit with its own specific innovations, but rather as a residual group which did not undergo changes characteristic of the languages of other groups. The genetic nature of the primary split between Formosan and other languages is not properly motivated, and thus is also questionable.

apply them. Without them a claim remains a hypothesis, which can not be fully justified.

Often, however, a genetic claim is based only on a list of comparisons, not supported by systematic phonological correspondences or tools to eliminate loans. In such cases formal justification of genetic nature of comparisons is substituted by the intuition of the linguist. This immediately puts the claim beyond formal comparative evaluation and, strictly speaking, it should be rejected, as not of a comparative nature: one cannot argue against other people's intuition.

If two forms are included in a comparison, it means that the linguist who proposed this comparison believes that those forms represent two separate, independent and uninterrupted developments of a single proto-form. If this belief can be confirmed with the help of systematic phonological correspondence, and arguments that the comparison does not result from against borrowing, then we are dealing with an etymology. Otherwise, we have a comparison of unknown genetic origin. Cognates in an etymology can be similar to each other, or they can be quite different. For example in the Sino-Tibetan etymology 'eight': Chinese *ba*, Zangskar dialect of Tibetan *yat*, Burmese *ṣi'*, Luchuan *ʔhen*<sup>55c</sup>. forms are quite different, while the meanings are identical. Transparent formal similarity between cognates is not however very important, when we are dealing with true etymologies: much more important is that it can be formally demonstrated that all such morphemes with different forms and meanings are various developments of a single proto-morpheme.

Working with languages which are not connected by a whole set of systematic phonological correspondences we do not have any formal means to prove that two morphemes should be included in a comparison. We can say only something like: 'Look, their forms and meanings are similar, so perhaps they can be traced back to a single source'. There is no way, however, to substantiate this suggestion. The danger of this situation, in the absence of any restrictions, is that we could bring together dissimilar forms, for example, Burmese and Yao morphemes 'fire': *mi*: and *tou*<sup>4</sup> (which are, in fact, not related) in a proto-form *\*toui* or *\*mitou* and use this ghost as evidence in a genetic claim. To avoid such mistakes we rather deal with comparisons in which the resemblances reveal phonologically transparent connections. For each comparison we should be able to work out a correlation between the syllabic structures of resemblances and a correlation of individual phonemes in these structures. If on the basis of Proto-Zhuang-Tai *\*phram*<sup>A</sup> 'hair' and Proto-Kam-Sui *\*pram*<sup>A</sup> 'hair' a proto-form *\*p-ram* 'hair' is suggested, I can not seriously argue against this pre-reconstruction, as it is based on clear similarity. But if this proto-form is connected with Proto AN *\*ra(m)but* 'hairy' via two intermediate stages like

*\*p-ram < \*[ra]p-ram[boc] = \*[ts,tʃ]-r-a(m)boc > \*ra(m)but*

(Benedict 1990: 204-205) I have the right not to believe in it: too many changes need to be proposed to justify this comparison, and none of them have any supporting evidence.

Meanings of resemblances should also correlate rather simply. Ideally they should be synonyms in a broad sense. No unusual correlations are permitted at this stage of investigation and I cannot accept such distant semantic connection as 'above' / 'north', 'accustomed' / 'friend, companion', 'father or grandparents' / 'the god, thunder', as proposed in the first three comparisons given in the Benedict's work on JAT (1990: 161-162). Semantic relations like

'ant' / 'ant, 'back of a blade' / 'back, ridge', 'hind part' / back, behind' (cf. the next three comparisons in Benedict 1990: 162) are more convincing.

The restrictions placed upon comparisons do not mean that I a priori reject all non trivial etymologies. What I am saying here is, that at the stage of collecting data (the heuristic level of justification of a genetic claim) we should try to avoid any cases which are not straightforward, as they can lead to wrong results. Only after a genetic claim is proven and phonological correspondences are established, can we deal with the more obscure cases. At this stage (in proper comparative research) they should not affect our conclusions about genetic relatedness of the languages.

Many genetic claims found in the literature are of a heuristic type, with some of them intuitively more acceptable than the other. What makes the difference? In all generally accepted cases linguists are dealing with Young and Developed families where similarities between the languages are transparent, and anyone (whether speaker or linguist) can detect them. This allows scholars to compile lists of comparisons, but without comparative phonology they can not detect loans and separate them from etymologies.

In some cases, however, a simple procedure can help to do this. It is based on the following considerations. At least six major groups of morphemes can be distinguished in a language's lexicon. The first group - *descriptive morphemes* - includes morphemes which represent sounds, or activities accompanied by sounds. Such morphemes have a relatively high chance of being sound symbolic. Formal similarities based on onomatopoeia, idiophony and other types of sound symbolism, do not indicate genetic relationship. At the same time descriptive morphemes are not necessary sound symbolic. Quite often, however, it is difficult or even impossible to judge whether a descriptive morpheme is symbolic. In some cases, historical phonology can be of assistance; in others, the question remains open. Given the high probability of sound symbolism in such cases, it is preferable not to include descriptive morphemes in comparisons at the heuristic stage of investigation.

The second lexical group includes so-called *cultural* morphemes, or lexical morphemes with meanings related to various cultural ideas. As it is quite common for people to borrow ideas together with the appropriate words, we can expect a certain proportion of borrowings among the cultural morphemes of a language.

The third group includes morphemes which belong to the so-called *core lexicon*. The meanings of such morphemes are universal and are represented in most languages of the world, so it is less likely that such morphemes would be borrowed between languages. Of course, borrowings in the core lexicon are known, but the chances of runing across them here are usually not as high as for those in the cultural lexicon. Sound symbolic morphemes are also less common among core morphemes.

It is not simple to define a list of meanings which should be included in the core lexicon. Such meanings are represented, for example, in the 100-item and 200-item lists used in lexico-statistics, but a more extensive list could also be suggested.

The fourth group is formed by *grammatical morphemes*. In principle, these morphemes can be either original or borrowed, but normally we expect that grammatical morphemes are resistant to borrowing. There are also less chances for such morphemes to be of the sound symbolic type. From this point of view



grammatical morphemes are similar to core morphemes, but unlike the latter they are not universal and in some languages, like Classical Chinese, grammatical morphemes are extremely rare.

The fifth group is represented by lexical morphemes which can be called *environmental*. Their meanings are associated with various natural, floristic and faunistic phenomena: names of different species of vegetation, animals, birds and so on. The origins of such morphemes in a language reflects the history of its speech community. If migrations occurred, we expect to find many borrowings among these morphemes. In other cases they may remained unchanged.

The rest of the morphemes of a language belong to the sixth group. The origin of its members is hard to predict: they can be borrowed, of sound symbolic nature or be retained from previous stages of the language development.

The six groups identified here are not mutually exclusive and the origin of a particular morpheme cannot be predicted simply by its group membership. This membership, however, indicates its probable development: a morpheme from a cultural group will more naturally be borrowed than for a morpheme from the core lexicon. This observation is used in a technique for primary evaluation of a genetic claim. If languages are transparently related, they always have a certain number of comparisons among morphemes from the core lexicon. For Old and Macro families such comparisons should be found between forms of the proto-languages under consideration. There are no commonly accepted language families with no core comparisons, and if such comparisons are not found, a genetic claim seems unreasonable, even if it is supported by comparisons based on grammatical and other types of morphemes. This assumption leads to the following semi-formal procedure, suggested more than 20 years ago in some talks given by Jakhontov, and presented here in a modified form.

A check of a genetic claim can be based on the same lists of morphemes as those used for lexicostatistics. Each list includes the main, semantically unmarked translations of the 100 core meanings found in a particular variant (dialect) of one of the languages under investigation. Comparing lists by studying entries with the same meanings, a linguist identifies comparisons, and separates them into original ones and loans. If the languages are transparently genetically related, they will always have a reasonable number of original comparisons. Without them a genetic claim is not valid.

Let us take as an example the relationship between three languages: Chinese, Tibetan and Burmese which belong to different branches of an Old Family, traditionally called Sino-Tibetan. Modern forms of these languages are so different that it is very difficult to detect similarities between Beijing's Mandarin, Lhasa Tibetan and Modern Spoken Burmese. An internal reconstruction of Chinese and evidence from the Tibetan and Burmese traditional orthography reduce the differences between the languages, and bring us to the level of a Developed Family (a situation similar to Indo-European with its archaic languages). At this level, similarities between the languages are more transparent, and a reasonable list can be collected. The main body of evidence that the three languages are genetically related is a list of several hundred comparisons (Peiros & Starostin 1966) which connect any pair or all three of these languages. They include lexical morphemes and pronouns, but comparisons of purely grammatical morphemes are not found. The number and

quality of comparisons rules out chance similarities as an explanation, leaving open only two possibilities: mass borrowing or genetic relationship.

Comparisons from the 100-item list indicate genetic relationship, as comparisons are found between any pair of these languages, as well as between all three of them:<sup>12</sup>

		Chinese	Tibetan	Burmese
1.	die	<i>siʃʔ</i>	<i>āčhi</i>	<i>siʃ</i>
2.	ear	<i>nhəʔ</i>	<i>rna</i>	<i>na;</i>
3.	fire	<i>smə:jʔ</i>	<i>me</i>	<i>mi:</i>
4.	fish	<i>ŋha</i>	<i>ña</i>	<i>ŋa;</i>
5.	kill	<i>sra:t</i>	<i>gsod</i>	<i>sat</i>
6.	long	<i>draŋ</i>	<i>riŋ</i>	<i>hrañ</i>
7.	name	<i>mheŋ</i>	<i>miŋ</i>	<i>ʔa.-mañ</i>
8.	short	<i>to:nʔ</i>	<i>thuŋ-thuŋ</i>	<i>tui</i>
9.	sun	<i>nit</i>	<i>ñi-ma</i>	<i>nij</i>
10.	two	<i>nij-s</i>	<i>gñis</i>	<i>hnac</i>

		Tibetan	Burmese
1.	black	<i>nag</i>	<i>nak</i>
2.	bone	<i>rus</i>	<i>ʔa.rui:</i>
3.	dog	<i>khi</i>	<i>khuj;</i>
4.	eat	<i>za</i>	<i>ca;</i>
5.	eye	<i>mjig</i>	<i>mjak.ci</i>
6.	hand	<i>lag</i>	<i>lak</i>
7.	heavy	<i>l<sup>h</sup>ʒid</i>	<i>lij;</i>
8.	know	<i>šes</i>	<i>si.</i>
9.	liver	<i>mčín</i>	<i>ʔa.sañ:</i>
10.	meat	<i>ša</i>	<i>ʔa.-sa:</i>
11.	moon	<i>zla</i>	<i>la.</i>
12.	nail	<i>sen-mo</i>	<i>lak-sañ:</i>
13.	near	<i>thag-ñe</i>	<i>ni:</i>
14.	neck	<i>mđriŋ</i>	<i>lañ-paŋ;</i>
15.	nose	<i>sna-khug</i>	<i>hna-khaŋ;</i>
16.	not	<i>ma</i>	<i>ma.</i>
17.	road	<i>lam-kha</i>	<i>lam:</i>
18.	salt	<i>chwa</i>	<i>cha;</i>
19.	snake	<i>sbrul</i>	<i>mruj</i>
20.	star	<i>kar-ma</i>	<i>kraj</i>
21.	tongue	<i>lče</i>	<i>hlja</i>
22.	tooth	<i>so</i>	<i>swa:</i>
23.	tree	<i>šij-sdoŋ</i>	<i>sac-paŋ</i>

		Chinese	Burmese
1.	dry	<i>ka:r</i>	<i>khrauk</i>
2.	horn	<i>kro:k</i>	<i>khjui</i>
3.	new	<i>sin</i>	<i>sac</i>
4.	night	<i>lia-s</i>	<i>ña.</i>
5.	sand	<i>sra:j</i>	<i>saj;</i>
6.	stone	<i>diak</i>	<i>kjauk</i>
7.	tail	<i>məjʔ</i>	<i>mri:</i>
8.	year	<i>nhí:n</i>	<i>hnac</i>
9.	yellow	<i>waŋ</i>	<i>wa</i>

<sup>12</sup> Starostin's Old Chinese reconstructions represent Mandarin words. Lhasa Tibetan and Standard Burmese forms are given in their traditional orthography. Most of the comparisons are well known (Shafer 1966; Benedict 1972; Peiros & Starostin ms).

		Chinese	Tibetan
1	I	<i>pha:j</i>	<i>pa</i>
2	louse	<i>srit</i>	<i>srig</i>
3	mouth	<i>kho:ʔ</i>	<i>kha</i>
4	this	<i>te</i>	<i>adi</i>
5	water	<i>tujʔ</i>	<i>chu</i>

Formal identification of the comparisons as etymologies is based on Sino-Tibetan comparative phonology as it is reconstructed in Peiros & Starostin 1996, but even without systematic correspondences the identity of the forms in most cases is quite obvious and is accepted by such linguists as Shafer, Benedict or Luce who worked without a complete set of phonological correspondences for these three languages.

Let us investigate what conclusions can be drawn from comparison of the Sino-Tibetan and Vietnamese 100-item lists.

In one comparison, the Vietnamese form is similar to those of all the other languages:

	Chinese	Tibetan	Burmese	Vietnamese
kill	<i>sra:t</i>	<i>gsod</i>	<i>sat</i>	<i>giết</i>

However, this comparison is not reliable. The Austroasiatic origin of the Vietnamese form is well known and the only formal similarity between the Vietnamese and those of other languages is the final *-t*.

No binary similarities between Tibetan and Vietnamese are found. Similarities between Burmese and Vietnamese, or Tibetan / Burmese and Vietnamese, are represented by one comparison each, remaining within the bounds of chance resemblance:

	Chinese	Tibetan	Burmese	Vietnamese
rain			<i>mui:</i>	<i>mua</i>
tongue		<i>lče</i>	<i>hlja</i>	<i>lưỡi</i>

The majority of comparisons include a Chinese form:

	Chinese	Tibetan	Burmese	Vietnamese
1 fly	<i>pəj</i>	<i>(āphir</i>	<i>pjam)</i>	<i>bay</i>
2 green	<i>che:ŋ</i>	<i>lʃ* aŋ-khu</i>		<i>xanh</i>
3 head	<i>s-luʔ</i>			<i>đầu</i>
4 heart	<i>səm</i>			<i>trái tim</i>
5 leaf	<i>lap</i>	<i>lo-ma</i>		<i>lá</i>
6 liver	<i>ka:n</i>			<i>gan</i>
7 near	<i>gənʔ</i>			<i>gần</i>
8 yellow	<i>waj</i>		<i>wa</i>	<i>vàng</i>

Even without any knowledge of the history of Southeast Asian languages we can suggest the only acceptable interpretation of these comparisons: they include chance similarities ('fly, 'leaf') and borrowings. As these comparisons are limited only to Chinese and Vietnamese and do not

include cases without Chinese, we can talk only about borrowings. The direction of borrowing (from Chinese to Vietnamese) is indicated by the fact that some Chinese forms ('green', 'yellow') in other Sino-Tibetan languages.

Let us now investigate a claim that Chinese is genetically related to Austronesian languages (Sagart 1994) using the same technique. Here one can find several types of comparison between the three Sino-Tibetan languages and Standard Malay. The first type is represented by Sino-Tibetan forms similar to Malay:

	Chinese	Tibetan	Burmese	Malay
1 die	<i>sij?</i>	<i>āchi</i>	<i>sij</i>	<i>mati</i>
2 dry	<i>ka:r</i>		<i>khrauk</i>	<i>kering</i>
3 long	<i>draŋ</i>	<i>riŋ</i>	<i>hrañ</i>	<i>pañjang</i>
4 road		<i>lam-kha</i>	<i>lam:</i>	<i>jalan</i>
5 sand	<i>sra:j</i>		<i>saj:</i>	<i>pasir</i>
6 tongue		<i>lče</i>	<i>hlja</i>	<i>lidah</i>

Most of these similarities are probably due to chance, but some may be loans. Comparisons which only hold between Tibetan and Malay are all of a chance nature:

	Tibetan	Malay
1 belly	<i>grod</i>	<i>perut</i>
2 sit	<i>bsdad</i>	<i>duduk</i>
3 stone	<i>rdo</i>	<i>batu</i>

No comparisons solely between Burmese and Malay are found. The comparisons with Chinese are more interesting:

	Chinese	Malay
1 cloud	<i>wən</i>	<i>awan</i>
2 egg	<i>ro:n?</i>	<i>telur</i>
3 foot	<i>kak</i>	<i>kaki</i>
4 hair	<i>pat</i>	<i>rambut</i>
5 root	<i>kə:r</i>	<i>akar</i>
6 salt	<i>lam</i>	<i>garam</i>
7 sleep	<i>duj</i>	<i>tidur</i>

Taken in isolation these could be treated as an indication of genetic relationship between Chinese and Malay. The addition of Burmese makes such a suggestion absolutely improbable. As it is clear that Chinese and Burmese are genetically related, one should expect to find comparisons between any these languages and Malay. The absence of reliable comparisons between Burmese and Malay leads us to interpret the data in exactly the same way as for the Sino-Tibetan languages and Vietnamese. The languages are unrelated, but there were some contacts between speakers of Chinese and Malay or of their ancestor languages.<sup>13</sup>

Now we can try to apply this procedure to the hypothesis that Austronesian languages are related to Kadai. Our Malay list reveals the similarities with the Siamese one:

<sup>13</sup>In fact the languages involved in the contacts were probably Proto-Chinese and a very ancient Austronesian language, possibly even Proto-Austronesian (Peiros & Starostin 1984, Peiros to appear).

		Malay	Siamese <sup>14</sup>	Proto-Kadai
1.	ashes	<i>abu</i>	<i>dau.B</i>	< *P- <i>dau</i> <sup>B</sup>
2.	black	<i>hitam</i>	? <i>dam.A</i>	< *? <i>nam</i> <sup>T</sup>
3.	die	<i>mati</i>	<i>ta:i.A</i>	< *I- <i>ta</i> (:) <i>i</i> <sup>A</sup>
4.	drink	<i>minum</i>	? <i>di:m.B</i>	
5.	dry	<i>kəriŋ</i>	<i>hə:ŋ.A</i>	
6.	eat	<i>makan</i>	<i>kin.A</i>	< * <i>kiVn</i> <sup>A</sup>
7.	eye	<i>mata</i>	<i>ta.A</i>	< *I- <i>nta</i> <sup>A</sup>
8.	fire	<i>api</i>	<i>vai.A</i>	< * <i>vVj</i> <sup>A</sup>
9.	green	<i>hijaw</i>	<i>khieu.A</i>	< *R- <i>m Vi</i> <sup>A</sup>
10.	know	<i>tahu</i>	<i>ru.C</i>	
11.	louse	<i>kutu</i>	<i>hau.A</i>	< * <i>trau</i> <sup>A</sup>
		(Malay)	Siamese	Proto-Kadai)
12.	moon	<i>bulan</i>	? <i>dian.A</i>	< *P-? <i>nian</i> <sup>A</sup>
13.	rain	<i>hujan</i>	<i>fon.A</i>	< * <i>vaN</i> <sup>A</sup>
14.	sand	<i>pasir</i>	<i>dra:i.A</i>	= Chinese borrowing
15.	this	<i>ini</i>	<i>ni.C</i>	
16.	tongue	<i>lidah</i>	<i>lin.C</i>	
17.	yellow	<i>kuning</i>	<i>hliəŋ.A</i>	< * <i>[C-]liəŋ</i> <sup>A</sup>

This set seems more convincing than the previous one and indicates that Siamese and Malay are probably genetically related, as suggested by the Kadai-Austronesian theory (part of JAT). There is still a possibility that borrowings could account for the comparisons (see Thurgood 1994 who, however, does not operate with a Proto Kadai reconstruction), but I know of no well-supported Kadai-Austronesian comparisons from the cultural lexicon<sup>15</sup> and the possibility of loans predominantly entering the core lexicon seems to me rather strange.

A study of Yao, a Miao-Yao language which Benedict also includes in his JAT family gives, a different picture. Eight possible comparisons are found between Yao and Siamese with no specific comparisons between Yao and Malay:

		Yao	Siamese	Malay
1	bird	<i>no.8</i>	<i>nok</i>	(+ AN etymology)
2	die	<i>taɪ.6</i>	<i>ta:i.A</i>	<i>mati</i>
3	egg	<i>tc!au.5</i>	<i>khai.B</i>	
4	fish	<i>bjau.4</i>	<i>pla.A</i>	
5	long	<i>da:u.3</i>	<i>ja:u.A</i>	
6	salt	<i>dzau.3</i>	<i>kliə.A</i>	
7	this	<i>na:i.3</i>	<i>ni.C</i>	<i>ini</i>
8	water	<i>wam.1</i>	<i>nam.C</i>	(+ AN etymology)

In two of these comparisons we also have resemblances from Malay. As no comparisons specific for Malay and Yao are known the data again can be

<sup>14</sup> The Siamese forms are given in transliteration. Proto-Kadai reconstructions are taken from Peiros, to appear.

<sup>15</sup> The whole list of comparisons which I can accept is included in Peiros to appear.

interpreted as an indication of contact between Siamese and Yao, but not as evidence for a direct genetic affiliation.<sup>16</sup>

Vietnamese and Khmer are genetically related: both belong to the Austroasiatic family. This fact is clearly confirmed by comparisons from their 100-item lists:

		Vietnamese	Khmer
1	bone	<i>xương</i>	<i>chəʔiŋ</i>
2	dog	<i>chó</i>	<i>chəkɛː</i>
3	earth	<i>đất</i>	<i>tiː</i>
4	foot	<i>chân</i>	<i>ʒaəŋ</i>
5	hair	<i>tóc</i>	<i>sok</i>
6	hand	<i>tay</i>	<i>taɟ</i>
7	horn	<i>sừng</i>	<i>səŋɛːŋ</i>
8	leaf	<i>lá</i>	<i>səlɪk</i>
9	louse	<i>chấy</i>	<i>caɟ</i>
10	meat	<i>thịt</i>	<i>sac</i>
11	neck	<i>cổ</i>	<i>koː</i>
12	new	<i>mới</i>	<i>thəmiː</i>
13	nose	<i>mũi</i>	<i>crəmuɰ</i>
14	one	<i>một</i>	<i>muəɟ</i>
16	root	<i>rễ</i>	<i>rik</i>
17	sand	<i>cát</i>	<i>khəsac</i>
18	sit	<i>ngồi</i>	<i>ʔəŋguɟ</i>
19	tail	<i>đuôi</i>	<i>kənduj</i>
20	this	<i>này</i>	<i>nəːh</i>
21	two	<i>hai</i>	<i>biːr</i>
22	water	<i>nước</i>	<i>dɪk</i>
23	what	<i>gì</i>	<i>səʔiː</i>
24	wind	<i>gió</i>	<i>khjaɭ</i>
25	year	<i>năm</i>	<i>chənam</i>

If we now compare Vietnamese, Khmer and Yao lists, the results confirm a hypothesis of their genetic relation. Here we find triples and binary comparisons between any pair of languages:

---

<sup>16</sup> It is possible that the Miao-Yao and Austro-Tai families are related, but to prove it one should look for comparisons between reconstructed 100-item lists for the corresponding proto-languages.

		Vietnamese	Khmer	Yao
1	bone	<i>xuong</i>	<i>chəʔiŋ</i>	<i>buŋ.3</i>
2	dog	<i>chó</i>	<i>chəkɛː</i>	<i>tcu.3</i>
3	horn	<i>sùng</i>	<i>sənɛːŋ</i>	<i>coŋ.1</i>
4	tail	<i>đuôi</i>	<i>kənduj</i>	<i>twei.3</i>
5	this	<i>này</i>	<i>nəːh</i>	<i>naːi.3</i>
6	two	<i>hai</i>	<i>biːr</i>	<i>i.1</i>
7	wind	<i>gió</i>	<i>khjal</i>	<i>dzjaːu.5</i>

		Vietnamese	Yao
1	cloud	<i>mây</i>	<i>mou.6</i>
2	come	<i>đi</i>	<i>taːi.2</i>
3	eye	<i>mắt</i>	<i>mwei.6-tsiːŋ.1</i>
4	long	<i>dài</i>	<i>daːu.3</i>
5	round	<i>tròn</i>	<i>tcʊn.2</i>
6	smoke	<i>khói</i>	<i>sjou.5</i>
7	you	<i>mày</i>	<i>mwei.2</i>

		Khmer	Yao
1	blood	<i>zhaːm</i>	<i>zjaːm.3</i>
2	rain	<i>bhliəŋ</i>	<i>bjuŋ.6</i>
3	tail	<i>kənduj</i>	<i>twei.3</i>

The same procedure can also be applied to the data Benedict uses to argue for a Japanese / Austro-Tai relationship. A comparison of the Proto-Japanese list with Austronesian (probably Proto-AN) reveals the following:

	Proto-Japanese <sup>17</sup>	Proto-AN
1. drink	<i>*nəm</i>	<i>*ʔinum</i>
2. eye	<i>*maiN</i>	<i>*mata</i>
3. fire	<i>*pə-i</i>	<i>*Capuy</i>
4. horn	<i>*tunwua</i>	<i>*tʊŋu</i>
5. tooth	<i>*pə</i>	<i>*Cipən</i>
6. tree	<i>*kəi</i>	<i>*kaSiw</i>
7. who	<i>*tə</i>	<i>*tʰa[y]i</i>
8. yellow	<i>*kui</i>	<i>*kuniŋ</i>

For of these Japanese forms (drink, eye, fire and tooth) have Altaic etymologies. Proto-Japanese reveals 25 comparisons with Korean and 15 with Tungusic (Starostin 1991:106) while neither Korean, nor Tungusic demonstrate any significant number of similarities with Proto Austronesian.

The result of applying this simple procedure to these languages suggests that we are dealing with four clear cut groups of them:

- (1) Sino-Tibetan: Tibetan, Burmese and Chinese
- (2) Austro-Tai: Siamese and Malay
- (3) Vietnamese, Khmer and Yao
- (4) Japanese (with other Altaic languages).

<sup>17</sup> Proto-Japanese forms and their Altaic etymologies are taken from Starostin 1991.

These conclusions, however, are preliminary and can be accepted only at the heuristic level of argumentation. It is worth remembering that the procedure cannot detect that groups (2) and (3) are possibly related, nor suggest any classification of languages within these groups.

The following considerations are important if the procedure is used:

1. Any two transparently related languages always show a certain number of comparisons from the 100-item list, usually more than 12-15. About 5 comparisons will usually be found between any two languages due simply to chance factors, and they do not indicate a genetic relationship. A lack of comparisons means that the languages cannot be directly connected to each other. To prove that they are remotely related one needs to study their proto-languages looking for comparisons between their reconstructed 100-items lists.

2. Any genetic claim based on 100-item lists should include comparisons from at least three languages: a binary comparison can lead to distorted results, as for the Chinese-Vietnamese relationship. A genetic claim not supported by a system of phonological correspondences should be based on interpretation of data from several languages, to aid in the detection of possible loans and other perturbations.

3. Comparative study of languages requires their systematic investigation: a comparable amount of data should be used and presented for each language involved. We could not take seriously a claim that languages A, B and C are genetically related which is based on twenty comparisons between languages A and B and on another twenty found in A and C, but not in B.

#### Phonological correspondences

Related languages always have comparisons involving forms from their core lexicon, usually supported by comparisons from other lexical groups. The presence of these comparisons, however, is not in itself enough to prove a genetic claim. Proof is possible only where a set of systematic phonological correspondences is presented. Without them, such a claim remains a more or less plausible hypothesis.

Phonological correspondences established from the whole collection of comparisons would be of two types:

- (i) those connecting phonemes of common origin, and
- (ii) those connecting phonemes in forms which are not genetically related (usually a result of borrowing).

A phonological correspondence which brings together reflexes of a particular proto phoneme or other features of a proto-language is called a *systematic correspondence*. Simply looking at a correspondence, however, we can never say if it is systematic: a reconstruction of the entire phonological system of a proto-language is needed before a correspondence can be identified with any certainty as being systematic.

A phonological correspondence supported in a sufficient number of comparisons is called a *regular correspondence*. The expression 'sufficient number of comparisons' is rather vague and usually depends on the number of comparisons found. If we have, for example, a thousand comparisons, a phonological correspondence supported by a hundred of them is regular, while a correspondence supported by only two comparisons is not. Quite often it is very difficult to decide if a correspondence is regular. What if a correspondence was based on two or three comparisons out of total a hundred reliable comparisons?



Regular correspondences can be found in comparisons which connect morphemes of common origin as well as borrowings. In cases of the latter type, such as between Chinese, Vietnamese and Japanese, regular correspondences occur only in situations of mass borrowings. Despite the fact that a system of regular phonological correspondences is known, these languages are not genetically related. Vietnamese and Japanese have intensively borrowed from Chinese during a rather limited period of time and the regularity of phonological correspondences reflects this fact.

What is necessary for proof of genetic relatedness, then, is a set of systematic (though not necessary regular) phonological correspondences. This set connects the phonological systems of the languages under investigation, which means that for each phoneme *a* of language *A* we have to find corresponding phonemes (sometimes  $\emptyset$ ) in all other languages. Usually, most of these systematic correspondences will be regular, supported by sufficient number of examples. A systematic correspondence can, however, be associated with a rare feature of the proto language and for this reason may be represented only in few etymologies. It is very important that the proposed set of systematic correspondences should connect all elements of forms included in a comparison, rather than being correct, say, only for initial consonant or tones.

Practically speaking, for any genetic claim we need to have tables of systematic phonological correspondences between all the languages discussed in the claim. Such tables should be given for all parts of their phonological systems, including syllabic structures, consonants (initial, medial, final as well as consonantal clusters), vowels and, if necessary, such suprasegmental elements as tones, registers or stress patterns. With the help of such tables we should be able to check whether the grouping of particular morphemes into etymologies is convincing or not. The regular correspondences in these tables should be identified and we can expect that they will be found in most etymologies.

For the genetic claims mentioned above, systematic phonological correspondences are given only for the SC theory. Here they connect only syllabic structures and consonants. Unfortunately, the published correspondences do not connect other parts of the phonological systems, primarily vowels, of the proto languages compared. Application of the correspondences to the forms included in comparisons shows that they are fairly consistent, and do not contradict each other. This means, at least for me, that it is highly probable that the Sino-Caucasian theory is correct, but the whole set of systematic correspondences is needed to provide the body of evidence formally required for the proof of the claim.

Systematic correspondences are not known for Austro-Tai and Miao-Austroasiatic hypotheses, which are supported by limited numbers of comparisons. As those comparisons can hardly be explained through borrowing, it is likely that further research would lead to the discovery of systematic correspondences among the proto-languages constituting each of these two families. The Sino-Austronesian and Japanese / Austro-Tai hypotheses are not supported by convincing comparisons, it is not surprising that sets of systematic phonological correspondences are not found. This means that both hypotheses should be rejected.

## References

- Benedict, Paul. K.  
 1942 Thai, Kadai and Indonesian: a new alignment in South Eastern Asia. - *American Anthropologist*, n. s. 44 (4): 576 - 601  
 1972 *Sino - Tibetan: A conspectus*. - Cambridge: Cambridge University Press.  
 1975 *Austro - Thai. Language and culture*. New Haven: Human Relations Areas Files Press.  
 1990 *Japanese / Austro-Tai* (Linguistica Extranea Studia 20). Ann Arbor: Karoma Publishers.
- Blust, Robert A.  
 1980 Austronesian etymologies. - *Oceanic Linguistics*, 19/1-2: 19 - 181  
 1983-84 Austronesian etymologies II. - *Oceanic Linguistics*, 22/23: 29 - 14  
 1986 Austronesian etymologies III. - *Oceanic Linguistics*, 25/1-2 : 1 - 123  
 1988 The Austronesian homeland: a linguistic perspective. *Asian Perspectives* 26: 45 - 67.  
 1989 Austronesian etymologies IV. - *Oceanic Linguistics*, vol. 28: 111 - 180
- Dempwolff, Otto  
 1934-38 *Vergleichende Lautlehre des Austronesischen Wortschatzes*. Berlin: Dietrich Reimer 3 vol.
- Jakhontov, Sergej E.  
 1980 Ocenka stepeni blizosti rodstvennyx jazykov. In V. N. Jarceva (ed.) *Teoreticheskie osnovy klassifikacii jazykov mira*. 148-157. Moscow: Nauka.  
 1981 Sootvetstija finalej v dialektax mjaio. *Vostokovedenie*, vol. 8: 63-72.
- Li, Fangkuei  
 1977 *A handbook of comparative Tai*. (Oceanic linguistics Special Publication 15) Honolulu: University Press of Hawaii.
- Lombard, Sylvia J.  
 1968 *Yao-English dictionary* (Data paper 69, Linguistics series 2). Ithaca: Cornell University Southeast Asia Program.
- Mahdi, Waruno  
 1994 Some Austronesian maverick protoforms with culture-historical implications. *Oceanic Linguistics* 33 1: 167-230; 2: 431-90.

- Martin, Samuel E  
1987 *Japanese language through time*. New Haven: Yale University Press.
- Matisoff, James A.  
1988 Proto-Hlai initials and tones: a first approximation. In J. A. Edmonson, D. B. Solnit *Comparative Kadai: linguistic studies beyond Tai*: 289 - 321. Summer Institute of Linguistics and The University of Texas at Arlington Publications in Linguistics 86.
- Nikolaev, Sergej L and Sergej A. Starostin  
1994 *A North Caucasian etymological dictionary*. Moscow: Asterisk Publishers
- Peiros, Ilia I  
to appear *Linguistic Prehistory in Southeast Asia*. Pacific Linguistics C-
- Peiros, Ilia I. and Sergej A. Starostin  
1984 Sino-Tibetan and Austro-Tai. *Computational analysis of Asian and African languages*. vol.22: 123-127.  
1996 *A Comparative Vocabulary of Five Sino-Tibetan Languages*. Fascicles 1-6. The University of Melbourne: Department of Linguistics.
- Purnell, Herbert C Jr.  
1970 *Toward a reconstruction of Proto-Miao-Yao*. PhD dissertation, Cornell University. Ann Arbor: University Microfilms International.
- Sagart, Laurent  
1993 Chinese and Austronesian: evidence for genetic relationship. *Journal of Chinese Linguistics* 21: 1- 62.  
1994 Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic Linguistics* 33/2: 271-308.
- Shafer, Robert  
1976 *Introduction to Sino-Tibetan*. Vol. 1. Wiesbaden: Otto Harrassowitz..
- Starostin, Sergej A.  
1982 Praenisejskaja rekonstrukcija i vneshnie svjazi enisejskix jazykov. In E. A. Alekseenko (ed.) *Ketskij Sbornik*. 144-237. Moscow: Nauka.  
1984 Gipoteza o geneticheskix svjazjax sinotibetskix jazykov s enisejskimi i severnokavkazskimi jazykami. *Linguisticheskaja rekonstrukcija i drevnejshaja istorija vostoka*. P.4: 19-38. Moscow: Nauka.

- 1989      *Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy.*  
Moscow: Nauka.
- 1991      *Altajskaja problema i proisxozhdenie japonskogo jazyka.*  
Moscow: Nauka.

Thurgood, Graham

- 1988      Notes on the reconstruction of Proto-Kam-Sui. In Jerold. A. Edmonson and David. B. Solnit *Comparative Kadai: linguistic studies beyond Tai.*: 179- 218. Summer Institute of Linguistics and The University of Texas at Arlington Publications in Linguistics 86.

Vine, Brent

- 1991      Indo-European and Nostratic. *Indogermanische Forschungen*. 96:9-35.

Vovin, Alexander

- 1994      Is Japanese related to Austronesian? *Oceanic Linguistics*, 33/2: 368-390.

Wang, Fushi

- 1985      *Sketch grammar of Miao* (in Chinese). Beijing: Chinese Academy of Social Sciences

# On Pronominal Systems

Richard A. Rhodes  
*University of California-Berkeley*

"It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbable, must be the truth."

Sherlock Holmes in *The Adventure of the Beryl Coronet*

**0. Introduction.**<sup>1</sup> The work of the historical linguist interested in establishing the common ancestry of two languages is like that of Sherlock Holmes himself. He or she must identify all the possibilities that might account for the systematic similarities between two languages (and any possible parallel deviations therefrom) and then eliminate all the possibilities except for that of common ancestry. This is as true for long-range comparativists attempting to argue for remote relationships between language families as it is for those working on relationships between languages within a family. But the layering of semantic shifts and the general replacement of vocabulary over time happens at a fast enough rate that relatively few vocabulary items show sufficient semantic stability over 10,000 year lengths of time to be reliably used in the comparative method. Because of this fact, it has been the practice of long-range comparativists to focus on certain core vocabulary items including, among other things, pronouns and pronominal agreement markers. This practice is generally asserted and not argued for, as in Shevoroshkin's (1990) introduction to *Proto-Languages and Proto-Cultures*:

"But as soon as I started to compare Salishan and Sino-Caucasian, I saw that practically all stablest roots (*pronouns* 'I', 'thou'; numerals 'two', 'three'; terms for body parts, etc.) show clear matching between Salishan and Sino-Caucasian (mostly between Salishan and North-Caucasian)" pg. 9 [emphasis mine, RAR]

The purpose of this paper is to raise a caveat regarding the use of pronouns and pronominal affixes in long-range comparison. I will bring out reasons for suspecting that pronominal morphology is not as stable as was previously assumed and that this instability renders it suspect as *prima facie*

<sup>1</sup>An earlier version of this paper was presented at a Linguistics Department Colloquium at the University of Hawaii-Manoa, March 5, 1996. I wish to thank David Stampe for his insightful comments on that presentation and I would like to give a special thanks to Johanna Nichols for sharing the database that she and David Peterson developed in connection with their 1996 paper.

evidence in long range comparison.

Those who assume the stability of pronominal morphology are assuming two things: a semantic stability and phonological stability, i.e., a resistance to all but *lautgesetzlich* change. The purpose of this paper is to examine these assumptions. It is not my intent to prove beyond a shadow of a doubt that pronominal roots are unstable *per se* but rather I will show that there is enough typological evidence to suggest rather strongly that some of the apparent phonological stability of pronominals is only apparent and that similarities among pronominal system might be due to factors which would undermine their usefulness in long range comparison.

I will not address the question of semantic stability at length, but I will point out that there exists a literature that calls the assumption of semantic stability in pronominal systems into doubt. In recent work by Helmbrecht (1996) numerous examples of pronouns and pronominals undergoing semantic change of person are cited. These include a variety of sources for first and second person singulars in languages of the North America. He found first person singular arising from first person plural in several Mayan languages, from second singular in Tsimshian, and from a deictic in Wintu. As for second singular, he shows that Tunica (and possibly Yuchi) second singular has a source in first singulars while in Tsimshian, Aztec, and Bella Coola the source in first plural. Helmbrecht's exceptions to the semantic stability of pronouns largely arise though various politeness strategies especially the replacement of second person pronominals by third person pronominals, as is widely attested in the history of various Indo-European languages.

In this paper I want to present evidence that calls into question the second assumption of stability—that pronominal morphology is sufficiently immune from all but *lautgesetzlich* phonological reshaping that it is a reliable source in long range comparison.<sup>2</sup> I will argue here that similarities between languages in the phonological content of pronominal morphology can exist for reasons other than genetic inheritance. To do so I will argue that there are principled factors which favor particular phonological shapes for members of small, syntactically coherent semantic domains, of which pronouns and pronominals are possibly the best example.

**1. Backgrounding and pronominals.** The thesis is that pronouns and pronominals show consistencies of syntax and usage across languages which affect the overall phonological shape of the system. In saying this I am not suggesting that there is iconic sound symbolism, i.e. [+nasal] = 'first person'. This position has been suggested, for example, in Gordon (1995) but is shown to be untenable by Nichols and Peterson (1996). Rather I am claiming that there are reasons that follow directly from the act of communication which explain why either *m* or *n* should appear in the vast majority of pronoun systems. These reasons have to do with the fact that in all languages, either pronouns or

<sup>2</sup>Morphological reshaping is, of course, as widespread in this core vocabulary as anywhere else in the vocabulary.

pronominals or both appear in prototypical usage as backgrounded in discourse structure and that they are therefore prototypically backgrounded phonologically. This very fact places significant and unignorable constraints on effective communication. Thus it dictates what the optimal phonological material will be out of which a pronoun/pronominal system can be made.

While the fact of the backgrounding is such that it cannot be ignored, there are a variety of effects that are found in various languages depending on other typological considerations in the construction of their pronoun and pronominal systems. Therefore we need to turn our attention to a brief overview of the typology of pronominals.

**1.1. Typological considerations in backgrounding.** Let us first explore the typology of pronominals with respect to the matter of backgrounding. There are two general types of languages as regards pronouns and backgrounding. The contrast arises between head marking and dependent marking languages (Nichols 1986). In dependent marking languages pronouns are frequently obligatory and occur backgrounded in the most prototypical uses as in 1a,b. In some such languages pronouns can even be backgrounded to the point of being omitted, as in 1c.—a fact often misinterpreted by syntacticians.

1. a. English  
I hit him. [aɪhɪtɪm]
- b. Choapan Zapotec  
Zionē. 'She went.' cf. zio nigula 'The woman went.'  
zio-nē  
went-she
- c. Vietnamese  
Đi đâu đó? 'Where are you going?'  
go where *emph*

In head marking languages, on the other hand, there tend to be two pronominal systems—one of full pronouns and one of pronominal affixes or clitics. In these languages the pronominals are used in normal contexts and the full pronouns are most commonly used in emphatic contexts. In such languages the typical arrangement is that the pronominal affixes are backgrounded and pronouns are not.

2. a. Southwestern Ojibwe
  - i. Ingii-waabamaa. 'I saw him.'  
nīn-gii-waabam-aa  
1ST-PAST-see-3RD (ANIM)
  - ii. Nīin go ingii-waabamaa. 'I saw him.'  
I *emph* I-saw-him

## 2. b. Sayula Popoluca

- i. Tinkáyw. 'I ate it'  
 t̥in-kay-w  
**1ST-ACTS-ON-3RD-eat-COMPLETIVE**
- ii. ʔi:ts t̥inkáyw. 'I ate it.'  
 I I-ate-it

In head marking languages the kinds of constraints on backgrounded material we will be discussing apply to pronominal affixes but not necessarily to independent pronouns. However, there is one common class of cases in which independent pronouns in head marking languages appear to be subject to the backgrounding constraints also. This arises in those head marking languages which have independent pronouns that are based on the possessive pronominal inflections of a pronoun root. Contrast the two languages cited in 2. In Ojibwe (as in all of Algonquian) the pronouns are inflected pronoun roots, as shown in 3. But in Sayula Popoluca (as in all of Mixe-Zoquean) the phonological content of the pronouns is independent of that of the pronominal inflection as shown in 4.

## 3. Southwestern Ojibwe

pronoun	possessed noun	transitive verb
niin	niiyaw	niwaabamaa
'I'	'my body'	'I see him/her'
giin	giiyaw	giwaabamaa
'you sg.'	'your sg. body'	'you sg. see him/her'
wiin	wiiyaw	owaabamaan
'he, she'	'his/her body'	'he/she sees him/her'

## 4. Sayula Popoluca

pronoun	possessed noun	transitive verb
ʔi:ts	t̥inwáy	t̥inʔéʔp
'I'	'my son'	'I see him/her'
mi:č	ʔinwáy	ʔinʔéʔp
'you sg.'	'your sg. son'	'you sg. see him/her'
heʔ	ʔiwáy	ʔiʔéʔp
'he, she'	'his/her son'	'he/she sees him/her'

In head marking languages with derived pronouns like Ojibwe the shape of the independent pronouns, which are not prototypically backgrounded as whole words, will nonetheless have properties that look like they are backgrounded because their shape follows from the shape of the pronominals which are backgrounded.

**1.2. The communicative problems posed by phonological backgrounding.** Having shown where in head and dependent marking languages to look for backgrounded pronouns and/or pronominal markers, we



turn to the question of what constraints backgrounding imposes on the phonology of such morphemes. There are three immediate problems that must be solved for phonologically backgrounded material to be communicatively effective:

- 1) *the identification problem*: one must be able to tell when one is hearing a morpheme of the relevant type,
- 2) *the differentiation problem*: one must be able to distinguish among the different morphemes of the type, and
- 3) *the pronunciation problem*: one must be able to pronounce the morphemes with relative lack of attention.

I will argue that these considerations are in partial conflict, and that they, therefore, stand in a dynamic tension which, depending on the way a particular language chooses to balance them, supports the overall shape of the system in that language. Crosslinguistically, these considerations define a range of optimal pronominal systems that can be verified by exploring the typology of pronominal systems. The next section of this paper will be devoted to exemplifying these principles referring to enough typology to show that genetic relatedness is not at work. I will work backwards through the three problems.

**2.1. The pronunciation problem.** In order to pronounce forms in the background ease of articulation is at a relatively high premium. The notion of ease of articulation is congruent to the notion of markedness. Thus a constraint to produce forms that are prototypically backgrounded will favor the occurrence of lesser marked segments. From this follows directly the widely observed property of pronominals (and function words in general) that they only very rarely contain highly marked segments and that they mostly contain relatively unmarked segments (e.g. Campbell, 1993, in the context of the long range comparison debate).

But an interesting study is reported in Gordon (1995). In this paper Gordon "does the math". He shows that in a genetically and areally balanced sample of 62 languages pronouns consist predominantly of the least marked segments.<sup>3</sup> He counts the percentage of languages which use each of the sounds. I repeat as 5. the top ten consonants and the top five vowels on his list (pg. 120, Table 2).

<sup>3</sup>Nichols and Peterson (1996) argue that Gordon's typological balance is somewhat flawed, which opinion I share. However, the complaint they raise does not significantly affect the point I am making here.

5.	phone	% of languages	phone	% of languages
	n	93.5	a	98.4
	m	75.5	i	90.3
	k	71.0	u	69.4
	t	67.7	o	56.5
	y	53.2	e	51.6
	w	43.5		
	h	40.3		
	ŋ	38.7		
	s	37.1		
	r	37.1		

While one might want to argue that there are problems with Gordon's work in detail, I, nonetheless, believe his conclusions excerpted in 5. are right. Furthermore, he argues that one cannot claim that pronouns use significantly restricted inventories because there are counterexamples in the form of languages with small phonological inventories that do not significantly restrict those inventories to form pronouns. If true, this only reinforces my position, that small inventories in pronoun/pronominal systems arise out of phonological markedness considerations alone, not out of a pressure of an unknown source which favors smaller inventories in pronoun/pronominal systems.

However, there is an inventory argument that can be made in at least some languages with small total phonological inventories based on the distributional frequency of the segments by token. Gordon's method simply left any such consideration out. For example, Ojibwe has 24 phonemes. They are shown in 6a. with the lenes, represented in the orthography by voiced symbols, unmarked.<sup>4</sup> Of these 16 are used in affixes and 13 are used in pronominal affixes, but it can be shown that the distributional frequency of segments in the 25 pronominal affixes are skewed both with respect to the frequency of phonemes in lexical morphemes and with respect to the frequency of phonemes in the full list of 46 affixal morphemes.<sup>5</sup> The 9 phonemes that are significantly above average are underlined in 6d. with the skewed percentages are in bold.

<sup>4</sup>This interpretation of data in language internal terms is what I suspect is lacking in Gordon's approach.

<sup>5</sup>This comparison of the list of pronominal affixes with the full list of affixes of which it is a proper subset is a hedge against the possibility that the small size of the affix lists will lead to a stochastic problem.

## 6. a. Ojibwe segment inventory

stops and affricates	lenis	b	d	ʃ	g	ʔ
	fortis	p	t	č	k	
fricatives	lenis		z	ʒ		
	fortis		s	š		
sonorants	nasals	m	n			
	glides	w		y		
vowels	long		short			
	ii	oo	i		o	
	e	aa			a	

b. Full affix list: aa (3 obj), ad (2>3 conj), ag (1>3 conj), ag (an pl), am (inan obj), an (inan pl), ban (pret), d/g (3 subj conj), dig (dub), en (dub), g (inan subj conj), g/gon (imp pl), gi (2 subj/poss), i (1 obj), ig (an pl), igo (inverse), igoo (pass), ik (3>2 conj), im (poss), in (2 obj), in (inan pl), ind (3 pass conj), ing (loc), ini (obv), ka (delayed imp), ke (neg imp), m (indef subj), min (1 pl), n (imp sg), n/naa (n registration), naan (1 pl), ni (1 subj/poss), o (3 subj/poss), oo (inan obj), sii (neg), sinoon (neg), w (irr), w (3 subj), waa (non-1 pl), yaan (1 sg subj conj), yaang (1 excl pl conj), yan (2 sg subj conj), yangid (1 pl incl > 3 conj), yangw (1 incl pl conj), yegw (2 pl conj), yok (pl imp)

c. Pronominal affixes: aa (3 obj), ad (2>3 conj), ag (1>3 conj), am (inan obj), d/g (3 subj conj), g (inan subj conj), gi (2 subj/poss), i (1 obj), ik (3>2 conj), in (2 obj), ind (3 pass conj), m (indef subj), min (1 pl), naan (1 pl), ni (1 subj/poss), o (3 subj/poss), oo (inan obj), w (3 subj), waa (non-1 pl), yaan (1 sg subj conj), yaang (1 excl pl conj), yan (2 sg subj conj), yangid (1 pl incl > 3 conj), yangw (1 incl pl conj), yegw (2 pl conj)

6.	d.	phoneme freq. in affixes	freq. in pron. affixes	freq. in lexical morphs
	<b>aa</b>	5%	<b>20 %</b>	6%
	<b>a</b>	9%	<b>24 %</b>	6%
	e	3%	4%	6%
	ii	1% (1)	-	3%
	<b>i</b>	<b>16 %</b>	<b>32 %</b>	8%
	oo	3%	4%	1%
	o	4%	4%	3%
	b	1% (1)	-	3%
	<b>d</b>	4%	<b>16 %</b>	3%
	j	-	-	1%
	<b>g</b>	<b>14 %</b>	<b>32 %</b>	5%
	p	-	-	1%
	t	-	-	1%
	ch	-	-	>1%
	k	4%	4%	4%
	h	-	-	1%
	<b>m</b>	4%	<b>12 %</b>	2%
	<b>n</b>	<b>21 %</b>	<b>40 %</b>	9%
	z	-	-	1%
	s	2% (2 allom)	-	3%
	zh	-	-	1%
	sh	-	-	3%
	<b>w</b>	4%	<b>12 %</b>	6%
	<b>y</b>	<b>6 %</b>	<b>24 %</b>	1%

The data in 6. show that the least marked segments are significantly more frequent in Ojibwe pronominal affixes, actually providing an argument for less marked segments playing a disproportionate role in pronominal affix formation, even in a language with a small inventory.

**2.2. The differentiation problem.** In order to produce forms that are easy for the hearer to distinguish when they are backgrounded one would optimally have forms that consist of phonological material of maximal acoustic distinctness. Thus one optimal arrangement would be to have an obstruent, a sonorant (especially a nasal), and a glide/glottal associated with each of the persons. If this is the correct way to look at things, then it shouldn't matter

what the place of articulation the obstruent, the sonorant/nasal, or the glide/glottal is. Nor should it matter which class of phoneme is associated with which person. In fact, by ignoring the point of articulation and paying attention only to the articulatory type, I can find in my database all of the six logical possibilities for matching obstruent, sonorant, and glide/glottal with person. Examples are given in 7.

7.	E. Ojibwe		Lai Chin		Yir-Yoront	
		Klamath		Mandarin		Hawaiian (pl.)
first person	n-	ni	ka-	wo	ŋoyo	ka-kou
second person	g-	ʔi	na-	ni	ŋoto	'ou-kou
third person	w-	bi	ʔa-	ta	ŋolo	la-kou

The evidence in 7. strongly suggests that the relation between person and articulatory type are arbitrary. From this we can deduce that there is pressure to shape pronoun/pronominal systems in such a way as to maximize the acoustic distinctness but clearly not arising from any sound symbolic link between one of the persons and one of the class of sounds. This fact is at the crux of my argument that pronoun/ pronominals cannot be take at face value in a determination of long-range relatedness.

**2.3. The identification problem.** The last problem for effective communication from material that is backgrounded is identifying a morpheme or morphemic complex as an instance of a particular syntactic type—in this case a pronoun/pronominal. This problem is different from the other problems in that it has solutions that involve factors other than just phonology. In particular, syntactic considerations can help in the identification of specific loci in the background which could be expected to contain pronominal information. But let us first look at languages which have phonological solutions to the identification problem. One kind of solution is to have an inflected pronominal root. The presence of the root signals the point at which the pronominal marking is located. Examples are readily found in many head marking languages, as exemplified in 8.

8. a. Lakhota (root = -iye)

	sg	pl	
first person	miye	ʔkiyepi	cf. mitake kj 'my older sister'
second person	niye	niyepi	cf. nitake kj 'my older sister'
third person	iye	iyepi	cf. takeku kj 'his older sister'

b. Tonkawa (root = -a--)

	sg	du	pl
first person	sa·ya	kewsa·ya	kewsa·ka
second person	na·ya	wena·ya	wena·ka
third person	ʔaye·la	ʔawe·la	ʔaye·ka

'... also'	sg	du/pl
first person	sa·x <sup>Wa</sup>	kewsa·x <sup>Wa</sup>
second person	na·x <sup>Wa</sup>	wena·x <sup>Wa</sup>
third person	?a·x <sup>Wa</sup>	?awax <sup>Wa</sup>
'by ...'s self'	sg	du/pl
first person	sa·cos	kewsa·cos
second person	na·cos	wena·cos
third person	?a·cos	?awacos

The other phonological strategy is to have a template which all (or most) pronoun/pronominals match. Examples of this sort are most frequently found in languages in Africa and Australia. Examples are given in 9.

9. a. Katla (Kordofanian)

	sg	pl
first person	ɲɔŋ	nɛn
second person	ɲaŋ	nɔn
third person	ɲuŋ	ɲin

template      C    V    C  
                 [+nas]   [+nas]

## 9. b. Yir-Yoront (Pama-Nyungan)

nominative	sg	du	pl
first person	(ŋ)oyo	ŋele (in) ŋelen (ex)	ŋopol (in) ŋeɣən (ex)
second person	(ŋ)oto, ŋoto	(ŋ)opol	(ŋ)epəl
third person	(ŋ)olo, ŋolo	pula	pilin

template (non-third person dual/plural)

**3. Typology of pronominals** To be able lay out our proposal fully it will be necessary to undertake a rather long digression to discuss three important pieces of typological background. Two of these regard pronoun systems *per se* and one regards affixal person marking systems. We will take them up in order.

**3.1 Pronoun systems.** Pronoun systems fall into a typology of three general types with respect to how third persons are handled.

**3.1.1 Full systems.** In the first type, which I will call the FULL

SYSTEM, there are true third person pronouns. They are clearly distinct from demonstratives in that the pronouns never modify nouns nor do they have any deictic implications.

10. a. English<sup>6</sup>

nominative system	sg	pl
first person	I	we
second person	you (arch thou)	you (arch ye)
third persons	he she it (≠ this, that)	they
oblique system	sg	pl
first person	me	us
second person	you (arch thee)	you
third persons	him her it (≠ this, that)	them

10. b. Chinese (Mandarin)

	sg	pl
first person	wo <sup>3</sup>	wo <sup>3</sup> men
second person	ni <sup>3</sup>	ni <sup>3</sup> men
third person	ta <sup>1</sup> (≠ zhe <sup>4</sup> , na <sup>4</sup> )	ta <sup>1</sup> men

In full systems the contrasts we will have to account for are (at least) the three-way opposition among all persons.

**3.1.2 Restricted systems.** The second type of pronoun system, which I will call a RESTRICTED SYSTEM, has no true third person pronouns. The demonstratives, which can both modify nouns and have deictic implications, are the only third person pronouns.

<sup>6</sup>In this and in following tables I will include both nominative and accusative forms since the systems are sometimes different. In all cases I will list the nominative first. If there are further distinctions among pronouns beyond person and number, I will separate sets within a person and separate them with a semi-colon, giving any further information necessary in parentheses.

## 11. a. Latin

nominative system	sg	pl
first person	egō	nōs
second person	tū	vōs
third person	is, ea, id (cf. is homō 'this man')	eī, eae, ea
oblique system	sg	pl
first person	mē	nōs
second person	tē	vōs
third person	eum, eam, id	eōs, eās, ea
b. Turkish		
	sg	pl
first person	ben	biz
second person	se	siz
third person	o	onlar
	(cf. o meşhur aktör 'that famous actor')	

In restricted systems the key contrast we will have to account for is the two-way contrast between first and second persons, because the shapes of the third persons are as much determined by the demands of the deictic systems in which they participate as by their participation in the pronoun system. This view is supported by the history of languages with restricted systems in which the third persons are generally unstable, e.g. the development of Latin's restricted system into systems in Romance languages which mostly have full systems.

**3.1.3 Mixed systems.** The third type of pronoun system, which I will call a MIXED SYSTEM, has a true third person pronouns, but the demonstratives, which can both modify nouns and have deictic implications, can also be used as third person pronouns without deictic implications. Two such systems are shown in 12.

## 12. a. Klamath

nominative	sg	pl
first person	ni	na-t
second person	ʔi	ʔa-t
third persons	bi	ba-t; sa
	ke· [= this];	ke-k sa [= these]
	ho-t [= that]	ho-t sa [= those]
	ne· [= that absent]	ne-k sa [= those absent]



accusative	sg	pl
first person	nis	na-ts
second person	mis	ma-ts
third persons	{ bas	mna-ls; sas
	{ ke-ks [= this];	ke-ýas [= these]
	{ honks [= that]	honky'as [= those]
	{ ne-ks [= that absent]	ne-ýas [= those absent]

## b. Southwestern Ojibwe

	sg	pl
first person	niin	niinawind (excl) giinawind (incl)
second person	giin	giinawaa
third persons	{ wiin	{ wiinawaa
	{ wa?aw (an.),	{ ongow (an.),
	{ o?ow (inan.) [=this]	{ onow (inan.) [=this]
	{ a?aw (an.,	{ ingiw (an.),
	{ i?iw (inan.) [=that]	{ iniw (inan.) [=that]

In analysis we will treat mixed systems like full systems with respect to the true third person pronominals, but leave the deictics out, because as can be easily shown deictics have separate analyses as a small systems whether they play a role in the pronominal system or not.

**3.2. Systems with syntactically optional pronouns.** At the outset we stated out that pronouns are prototypically backgrounded. But on closer scrutiny this turns out not to be entirely true. There are languages in which pronouns are syntactically optional. For the purposes of this paper we need to explore the status of pronouns in such languages. Sidestepping the most frequently asked question about such languages, viz. the syntactic one of whether in such languages affixes marking person and/or number represent agreements or whether they are pronominal arguments, the crucial question for our purposes here is: In such languages are independent pronouns emphatic (and therefore not prototypically backgrounded) or are they obligatory (and therefore prototypically backgrounded)? This question affects our analysis of pronoun systems but not that of the pronominal affix systems, which are always backgrounded. For convenience in talking about the class of languages with optional pronouns, I will call all such languages pro-drop, but I do so without any intent to commit myself to defending that syntactic position for any specific language.<sup>7</sup>

<sup>7</sup>In fact I believe there are two types of pro-drop languages. Head-marking pro-drop languages are systematically pronominal argument languages (typical of the New World) and dependent-marking pro-drop languages which are truly pro-drop (typical of East Asia).

Typically pronouns in pro-drop languages are emphatic, i.e. not prototypically backgrounded. As a result they are frequently "long", i.e. four or more segments long for singulars and in hard-core head marking languages—those which mark two arguments in the verbs and inflect for possessor in nouns—it is common for the pronouns to be analyzable into a possessed pronominal stem as in the Lakhota forms cited in 8a. above and often to show a fuller paradigm including adverbial extensions as in the Tonkawa examples cited in 8b. above.

Since the full forms of pronouns in this language type are not backgrounded, we can safely set them aside as not belonging to the class of pronoun systems characterizable by the analysis we are proposing, although those that are amenable to an internal morphological analysis thereby answer the demands of the identification problem by being internally consistent. Furthermore, the person/number part of morphologically analyzable pronouns will be covered by an analysis of personal affixes.

**3.3. Systems with distinct oblique subsystems.** Many languages that mark case on nominals show distinct subsystems for oblique forms of pronouns. Examples are found above in Latin in 11a., and in Klamath in 9a. For our purposes we can treat such subsystems as distinct because they contrast within a single syntactic slot and the thrust of our inquiry is to lay out the phonological logic of entities in direct contrast. To the extent that we can give a comprehensive analysis, so much the better, but the approach I am taking here does not require us to do so.

**3.4 Pronominal affix systems.** The morphological typology of pronominal affix systems is much more complex. I will not be able to give a thoroughgoing typology here. For our purposes we need to distinguish pronominal systems along three parameters: 1) the number of arguments represented, 2) the number of person/number distinctions made, and 3) the amount of systematic category conflation within the system.

**3.4.1 Number of arguments.** There are systems which agree with one argument, those which agree with two arguments, and those which agree with three arguments. Systems of the latter two types are diagnostic of head-marking languages. Languages which mark a single argument differ in which argument they mark depending on the syntactic typology to which they belong. If the agreement pattern is nominative-accusative, the overwhelming majority of languages marking one argument mark subjects, if the agreement pattern is ergative-absolutive, then the majority mark absolutive. Such languages as so common as not to need exemplification. Languages which mark two arguments are in four basic types: nominative-accusative, stative-active, ergative-absolutive, and inverse. Examples are given in 13.

13. a. nominative-accusative system (singulars only)  
Classical Nahuatl (subject underlined, object bold)

noca	'call'			
	<i>intrans</i>	'... <b>me</b> '	'... <b>you</b> '	'... <b>him</b> '
' <u>I</u> ...'	<u>n</u> inoca	—	<u>n</u> imicnoca	<u>n</u> iknoca
' <u>you</u> ...'	<u>t</u> inoca	<u>t</u> in <b>e</b> c'noca	—	<u>t</u> iknoca
' <u>he</u> ...'	noca	<b>n</b> ec'noca	<b>m</b> icnoca	<b>k</b> inoca

- b. stative-active system (singulars only)  
Lakhota (Siouan) (actives underlined, statives bold)

kakiže 'suffer' (stative intransitive)

	' <u>I</u> ...'	' <u>you</u> ...'	' <u>he</u> ...'
<i>stat.-intrans</i> :	<b>m</b> akakiže	<b>n</b> icakiže	kakiže

kaštaka 'strike' (active intransitive and transitive)

	<i>act.-intrans</i>	'... <b>me</b> '	'... <b>you</b> '	'... <b>him</b> '
' <u>I</u> ...'	<u>w</u> akaštaka	—	<u>c</u> icaštaka	<u>w</u> akaštaka
' <u>you</u> ...'	<u>y</u> akaštaka	<b>m</b> aya <b>k</b> aštaka	—	<u>y</u> akaštaka
' <u>he</u> ...'	kaštaka	<b>m</b> akaštaka	<b>n</b> icaštaka	kaštaka

13. c. absolutive-ergative system  
Tzutujil (Mayan) (absolutives underlined, ergatives bold)

-wari 'sleep'

	' <u>I</u> ...'	' <u>you</u> ...'	' <u>he</u> ...'
<i>intrans.</i> :	<u>š</u> <u>i</u> n <b>w</b> ari	<u>š</u> <u>a</u> t <b>w</b> ari	š <b>w</b> ari

-čey 'hit' trans:

	' <u>I</u> ...'	' <u>you</u> ...'	' <u>he</u> ...'
' <u>I</u> ...'	—	<u>š</u> <u>a</u> t <b>n</b> u <u>č</u> ey	<u>š</u> <u>i</u> n <u>č</u> ey
' <u>you</u> ...'	<u>š</u> <u>i</u> n <b>a</b> a <u>č</u> ey	—	<u>š</u> <u>a</u> a <u>č</u> ey
' <u>he</u> ...'	<u>š</u> <u>i</u> n <b>r</b> u <u>č</u> ey	<u>š</u> <u>a</u> t <b>r</b> u <u>č</u> ey	<u>š</u> <u>u</u> <u>č</u> ey

## d. inverse system (singulars only)

Plains Cree (Algonquian) (subject underlined, object bold, the doubly underlined morphology indicates whether to interpret the person affixes as subject or object)

wâpi-/wâpam- 'see'				
<i>intrans</i>		'... you'	'... me'	'... him'
'you ...'	<u>kiwâpin</u>	—	<u>kiwâpam<sub>in</sub></u>	<u>kiwâpamâw</u>
'I ...'	<u>niwâpin</u>	<u>kiwâpamitin</u>	—	<u>niwâpamâw</u>
'he ...'	<u>wâpiw</u>	<u>kiwâpamik</u>	<u>niwâpamik</u>	{ <u>wâpamêw</u> <u>wâpamik</u> }

From our point of view what is important is whether the affixal subsystems marking persons in a particular grammatical relation have systematic zeroes or not. Such zeroes are found in the examples in 13. in the Nahuatl subject system in 13a., in both the Lakota subject and object system 13b., and in the Tzutujil absolutive system in 13c. In the overwhelming majority of cases, the zero is third person. Pronominal affix systems with systematic zeroes are to be analyzed as containing only a two-way contrast, following the same reasoning as that for pronoun systems with only deictic third persons.

**3.4.2 Number of contrasts.** There are three types of languages with respect to person contrasts. One (rather rare) type has affixes distinguishing just one person. Chitimacha agreement, for example, distinguishes only first person vs. non-first person. More commonly affix systems make three distinctions in person, although it is reasonably frequent that in systems with category conflation the number of person distinctions is fewer in more semantically complex points of the paradigm. For example, in German the plural agreements show fewer distinctions than the singular, thus *-en* marks [non-third plural], as shown in 14.

## 14. German

'go'	sg	pl
first person	<i>gehe</i>	<i>gehen</i>
second person	<i>gehst</i>	<i>geht</i>
third person	<i>geht</i>	<i>gehen</i>

This is of no particular consequence to us other than that it might affect the overall shape of a solution by changing the number of contrastive points.

**3.4.3 Category conflation.** Affixes marking person are very frequently conflated with other categories of verb inflection systematically—most commonly with number, but also with tense/aspect, subordinating morphology, and the like. This phenomenon is so common as not to warrant exemplification. But there are also systems with systematically split confluents. For example, Algonquian languages in that variety of subordinate inflection

called *conjunct* by Algonquianists, conflate subordination, person, gender, and number in non-third persons, but in third person number is a separate morpheme, as illustrated by the Southwest Ojibwe example in 15.

15. Ojibwe

'go home' <i>conjunct</i>	sg	pl
first person	giweyaan	{ giweyang (incl) giweyaang (excl)
second person	giweyan	giweyeg
third person	giwed	giwewaad

The importance of the fact of conflation to us is that it means that the system must be subjected to our analysis as involving as many distinct unanalyzable morphemes as there are in the system. Thus the German system presented in 14. has a five-way opposition (or possibly a four-way opposition, depending on the categorization system) and the Ojibwe system shows a six-way opposition.

**4. Semantic preliminaries.** Now that we have laid out a typology of pronouns and pronominal affixes which allows us to recognize which the relevant contrasts are whose phonological shape might be susceptible to our analysis, we need to talk about the relevant categories and our representation of them. We will propose that the semantic categorization of pronouns is, as we noted at the beginning of this paper, "clean", i.e. the basic person distinctions, speech act participant vs. non-speech act participant, and speaker vs. hearer, are not of the kind that require us to apply complex categorization notions like radial category. So we will propose the featural shorthand in 16a. to account for the semantics of person. Similarly we propose the featural shorthand in 16b. to account for the semantics of number.<sup>8</sup>

16. a. person

[± speaker]	'speaker' vs. 'non-speaker' (= ± first person)
[± hearer]	'hearer' vs. 'non-hearer' (= ± second person)
[± SAP]	'speech act participant' vs. 'non-speech act participant' (= ± third person)

<sup>8</sup>Because of its rarity I have left out trial/paucal. One might argue that number, especially trial/paucal, is semantically messy. Furthermore, the distinctions group plural vs. individuated plural are easy to find and I have provided no mechanism for accounting for such distinctions. But to the best of my knowledge, while these latter more semantically complex distinctions do occasionally appear in verb inflection, they are not ever conflated with person.

## b. number

[± pl]	'one' vs. 'more than one' (= ± plural)
[± dual]	'two' vs. 'not two' (= ± dual)

It may seem redundant at first that we propose more features than logically necessary, but, being a naturalist, I maintain that these categories are both natural (arising from cognitive considerations) and substantive and therefore it does not "cost" the theory anything to have more than the logically minimum number.

The features in 16. readily provide all the contrasts necessary. The one necessary person distinction that this notation readily makes which may not be immediately obvious is that of inclusive/exclusive first person. By using [± hearer] independently of [± speaker] it is possible to indicate this distinction, as in 17a. Such an analysis receives support from affix systems like those found in Algonquian languages which have affixes that mean [+ hearer] regardless of the value of [speaker] as shown in 17b. (The relevant morphemes are glossed in 17c.)

$$17. \quad a. \quad \text{exclusive} = \begin{bmatrix} + \text{ speaker} \\ - \text{ hearer} \end{bmatrix} \quad \text{inclusive} = \begin{bmatrix} + \text{ speaker} \\ + \text{ hearer} \end{bmatrix}$$

## b. Meskwaaki (Fox) (Algonquian)

'board (a vehicle)'	sg	pl
first person	nepo·si	{ kepo·sip <u>ena</u> (incl) nepo·sip <u>ena</u> (excl)
second person	kepo·si	kepo·sip <u>wa</u>
third person	po·si <u>wa</u>	po·si <u>waki</u>

## 17. c. glosses of the relevant morphemes

person markers:	ke- = [+ hearer]	ne- = $\begin{bmatrix} + \text{ speaker} \\ - \text{ hearer} \end{bmatrix}$
plural markers:	-p <u>wa</u> = $\begin{bmatrix} - \text{ speaker} \\ + \text{ hearer} \\ + \text{ pl} \end{bmatrix}$	-p <u>ena</u> = $\begin{bmatrix} + \text{ speaker} \\ + \text{ pl} \end{bmatrix}$

Let me briefly argue for the number features in 16. They are set up such that [+dual] implies [+plural]. In part this represents a number markedness hierarchy like that in 18a., but more cogently it reflects the fact that some languages mark duals in such a way that reflects the fact that they are also plural. An example is given in 18b. with the sketch of a morphological analysis in 18c.

18. a. singular > plural > dual (> trial/paucal)

b. Yawlamni (Yokuts)

accusative	sg	dual	pl
first person	nan	{ na-nikwa (ex) makwa (in)	na-ninwa (ex) maywa (in)
second person	mam	ma-mikwa	ma-minwa
third person	?amam	?ama-mikwa	?ama-minwa

c. glosses of the relevant morphemes

persons: n- = \b\bc\[(\a\al(+ speaker)) m- = \b\bc\[(\a\al(- speaker))  
?a- = [- SAP]

root: -a- = 'personal pronoun'

case: -n ~ -m ~ -Ø =  $\begin{bmatrix} - \text{nom} \\ - \text{poss} \end{bmatrix}$ <sup>9</sup>

plural markers: -ik- =  $\begin{bmatrix} + \text{pl} \\ + \text{dual} \end{bmatrix}$  -in- ~ -y =  $\begin{bmatrix} + \text{pl} \\ - \text{dual} \end{bmatrix}$

-wa- = [+ pl]

**5. Analyses.** In this section I will lay out analyses for several pronoun and pronominal systems using the approach that I have just outlined.

**5.1. Distinctness preference systems.** Let me start with systems that radically favor the distinctness preference over system coherence.

**5.1.1 Mandarin pronouns.** The pronouns of Mandarin are laid out in 10b. above. The number is straightforwardly susceptible to a conventional morphemic analysis, so it is irrelevant to the current exploration. But the persons are not amenable any type of conventional analysis. By eyeballing the three person markers we can see that both the consonantism and the vocalism are well spread in phonological space. The vowels are a fairly good set for distinguishing labiality (*o*) vs. palatality (*i*) vs. sonorance (*a*). Similarly the consonants are so well spread that they can be distinguished solely in terms of major class features, being a glide (*w*) vs. a nasal (*n*) vs. an obstruent (*t*). A conventional sound symbolic approach gives the results in 19. connecting phonology directly to semantics. Notice that because we are saying that the system in 19. is a radical distinctness favoring system, the semantic-phonological link of interest is only for parsing, i.e. the hearer, having, by whatever means, determined that the morpheme in question represents a pronoun, uses the semantic-phonological equations in 19. to determine the person. Put another way, the hearer need know only that the morpheme in question is a

<sup>9</sup>The gloss on this morpheme reflects the fact that this morpheme also occurs in the dative, ablative, and locative. The length in the plural marker -wa- is only supported in these other oblique forms.

pronoun and then hearing obstruence knows immediately that it is a third person, he need not hear for sure that it is exactly an apical or that it is an aspirated stop. The mere obstruence is enough to make it possible to identify the person. This is optimal design for picking semantics out of backgrounded material. The same type of logic applies to the connection between nasality and second person, and semivocality and first person.

19. Chinese (Mandarin)

a. Consonantism

$[\alpha \text{ cons}] = [-\alpha \text{ speaker}]$

$[\alpha \text{ son}] = [\alpha \text{ SAP}]$

pronouns	person features	consonants	forms
first person	$\begin{bmatrix} +\text{speaker} \\ +\text{SAP} \end{bmatrix}$	$\begin{bmatrix} -\text{cons} \\ +\text{son} \end{bmatrix}$	$wo^3$
second person	$\begin{bmatrix} -\text{speaker} \\ +\text{SAP} \end{bmatrix}$	$\begin{bmatrix} +\text{cons} \\ +\text{son} \end{bmatrix}$	$ni^3$
third person	$\begin{bmatrix} -\text{speaker} \\ -\text{SAP} \end{bmatrix}$	$\begin{bmatrix} +\text{cons} \\ -\text{son} \end{bmatrix}$	$ta^1$

19. b. Vocalism

$[\alpha \text{ rd}] = [\alpha \text{ speaker}]$

$\begin{bmatrix} -\text{back} \\ \alpha \text{ ft} \end{bmatrix} = [\alpha \text{ hearer}]$

pronouns	person features	vowels	forms
first person	$[+\text{speaker}]$	$\begin{bmatrix} +\text{bk} \\ -\text{ft} \\ +\text{rd} \end{bmatrix}$	$wo^3$
second person	$\begin{bmatrix} -\text{speaker} \\ +\text{hearer} \end{bmatrix}$	$\begin{bmatrix} -\text{bk} \\ +\text{ft} \\ -\text{rd} \end{bmatrix}$	$ni^3$
third person	$\begin{bmatrix} -\text{speaker} \\ -\text{hearer} \end{bmatrix}$	$\begin{bmatrix} -\text{bk} \\ -\text{ft} \\ -\text{rd} \end{bmatrix}$	$ta^1$

**5.1.3 Ojibwe person affixes.** The person affixes of Ojibwe are part of the pronoun system which is given in 20. (repeating the relevant portion of 12b. above).



## 20. Southwest Ojibwe

	sg	pl
first person	niin	niinawind (excl) giinawind (incl)
second person	giin	giinawaa
third person	wiin	wiinawaa

The analysis of these forms is similar to that of the Mandarin forms in 19., except that a basic morphemic analysis is necessary first.

## 21. a. glosses of the relevant morphemes

persons:	$n- = \begin{bmatrix} + \text{ speaker} \\ - \text{ hearer} \end{bmatrix}$	$g- = [+ \text{ hearer}]$
	$w- = [- \text{ SAP}]$	
root:	$-iina = \text{'personal pronoun'}$	
plural markers:	$-wind = \begin{bmatrix} + \text{ speaker} \\ + \text{ pl} \end{bmatrix}$	$-waa = \begin{bmatrix} - \text{ speaker} \\ + \text{ pl} \end{bmatrix}$

## 21. b. Consonantism

$[\alpha \text{ cons}] = [\alpha \text{ SAP}]$
$[\alpha \text{ son}] = [-\alpha \text{ hearer}]$
$[+nas] = [+speaker]$

person features	consonants	forms
$\begin{bmatrix} + \text{ speaker} \\ - \text{ hearer} \\ (+ \text{ SAP}) \end{bmatrix}$	$\begin{bmatrix} + \text{ cons} \\ + \text{ son} \\ + \text{ nas} \end{bmatrix}$	$n-$
$\begin{bmatrix} + \text{ hearer} \\ (+ \text{ SAP}) \end{bmatrix}$	$\begin{bmatrix} + \text{ cons} \\ - \text{ son} \\ - \text{ nas} \end{bmatrix}$	$g-$
$\begin{bmatrix} (- \text{ hearer}) \\ - \text{ SAP} \end{bmatrix}$	$\begin{bmatrix} - \text{ cons} \\ + \text{ son} \\ - \text{ nas} \end{bmatrix}$	$w-$

**5.1.4. Latin subject agreement.** Let me conclude the discussion of distinctness preference systems with a quick look at a relatively complicated case. The person agreement markers of Latin are given in 22. Here there is a systematic conflation of person and number (at least synchronically) yielding a more complex system having six points rather than three, and there are two subsystems one of which is specialized for the perfect.

## 22. Latin

non-perf.	sg	pl
first person	-ō ~ -m	-(i/u)mus
second person	-s	-tis
third person	-t	-unt ~ -int
perf.	sg	pl
first person	-ī	-(i)mus
second person	-stī	-stis
third person	-t	-ērunt

As a result the analysis is more complex. It is outlined in 23.

## 23. a. The more marked the semantic contents, the relatively longer the form. (Panini's Law)

- 1) Singulars have one less vowel and one less consonant than non-singulars.
- 2) The perfect markers are as long or longer than their non-perfect counterparts.

## 23. b. Core analysis:

$\begin{bmatrix} +\text{obs} \\ -\text{cont} \end{bmatrix}$	=	[-speaker] -t, -unt ~ -int, -ērunt, -tis, -stī, -stis
$\begin{bmatrix} +\text{obs} \\ +\text{cont} \end{bmatrix}$	=	[+SAP] -s, -tis, -stī, -stis, -(i)mus
$\begin{bmatrix} -\text{obs} \\ +\text{lab} \end{bmatrix}$	=	[-hearer] -ō ~ -m, -(i)mus, -unt, -ērunt

Notice that as the system gets more complex the parsing templates become less specific and thus tell relatively less, as we can see in 24. Even so they still may tell a lot. Applying them to each of the points in the Latin paradigm, fairly good coverage is achieved.

24.	-ō	=	$\begin{bmatrix} -\text{obs} \\ +\text{lab} \end{bmatrix}$	=	[-hearer]
	-s	=	$\begin{bmatrix} +\text{obs} \\ +\text{cont} \end{bmatrix}$	=	[+SAP]
	-t	=	$\begin{bmatrix} +\text{obs} \\ -\text{cont} \end{bmatrix}$	=	[-speaker]
	-(i)mus	=	$\begin{bmatrix} -\text{obs} \\ +\text{lab} \end{bmatrix}, \begin{bmatrix} +\text{obs} \\ +\text{cont} \end{bmatrix}$	=	$\begin{bmatrix} -\text{hearer} \\ +\text{SAP} \end{bmatrix}$
	-tis	=	$\begin{bmatrix} +\text{obs} \\ -\text{cont} \end{bmatrix}, \begin{bmatrix} +\text{obs} \\ +\text{cont} \end{bmatrix}$	=	$\begin{bmatrix} -\text{speaker} \\ +\text{SAP} \end{bmatrix}$
	-unt	=	$\begin{bmatrix} -\text{obs} \\ +\text{lab} \end{bmatrix}, \begin{bmatrix} +\text{obs} \\ -\text{cont} \end{bmatrix}$	=	$\begin{bmatrix} -\text{hearer} \\ -\text{speaker} \end{bmatrix}$

It's probable that one should supplement the system in 23. with ad hoc clauses, e.g. [+long] = [+speaker], but the metarules for doing this kind of analysis are not yet sufficiently clear to know how to do it consistently. Since this paper is intended to be programmatic, providing an outline of the kind of analysis should be sufficient and we can move on.

**5.2. System coherence systems.** Let us turn to systems that favor system coherence over the distinctness preference. In order to say that a system favors system coherence we have to be able to abstract a template that is sufficiently contentful that it matches only a small percentage of the morphological lexicon. The Katla system in 9a. above is an example. I also want to include systems in which one point out of six or so doesn't fit. Thus I would want to count the Yir-Yoront (Pama-Nyungan) system (cited in 9b. and repeated below as 25.) as an example that strongly favors system coherence in spite of the fact that it contains forms of in the third dual and plural which do not match the template.

25. a. Yir-Yoront (Pama-Nyungan)

nominative	sg	du	pl
first person	(ŋ)oyo	ŋele (in) ŋelen (ex)	ŋopol (in) ŋeŋan (ex)
second person	(ŋ)oŋo, ŋoŋo	(ŋ)opol	(ŋ)epəl
third person	(ŋ)olo, ŋolo	pula	pilin

	C	V	C	V	(C)
b. template	[+nas]	$\begin{bmatrix} -\text{hi} \\ -\text{lo} \end{bmatrix}$	$\begin{bmatrix} -\text{hi} \\ -\text{lo} \end{bmatrix}$		

The template in 25b. has the second consonant as the least specified obligatory consonant and therefore the locus most capable of bearing contrast. The

template, though not characterizable as a morpheme under traditional analysis, can be argued to have a status independent of this consonant as shown by the clitic forms of the singulars given in 26.

26. Yir-Yoront (Pama-Nyungan)

nominative sg	full	clitic
first pers	(ŋ)oyo	y
second pers	(ŋ)oto, ŋoto	r <sup>10</sup>
third pers	(ŋ)olo, ŋolo	l

It is probably also the case that pronominal systems that have a ready morphological analysis should count as instances of weighting system coherence highly. This would include systems like the Ojibwe system in 20. above, and those that also favor coherence in the person affixes like the Lakohta given in 8a. and repeated here in a fuller paradigm in 27. in which all the non-zero prefixes are [+nasal].

27. Lakhota

	sg	dual	pl
first person	<i>miye</i>	<i>ykiye</i>	<i>ykiyepi</i>
second person	<i>niye</i>		<i>niyepi</i>
third person	<i>iye</i>		<i>iyepi</i>

If analyzable systems count as favoring coherence then systems like the Ojibwe system are, in fact, balanced between the demands of distinctness and coherence.

**5.3. Templates as emergent.** Let me conclude this part of the paper with an observation that I believe that the templates are emergent forms in the grammar. I like to make a brief suggestion as to why that would make sense. In recent work, (Rhodes, 1996) I claim that in the semantic field of breaking and tearing, there is an emergent template that shapes analogical language change in some languages of the Algonquian language family. The facts for Ojibwe are summarized in 28.

28. a. Proto-Algonquian

PA \*po-θk(w)- 'broken/torn off/apart'

PA \*ta-tw- 'torn open (of flexible objects)'

PA \*pa-šk- 'burst/crack open (of rigid objects)'

b. emergent template *p* V ([+cont]) *k* (w)

<sup>10</sup>The relation between *t* and *r* in the relevant contexts is morphophonemically regular.

## 28. c. Common Ojibwe reflexes

poohkw-	'broken (of stick-like objects)'
pahk-	'broken (of string-like objects)'
piikw-	'torn (of sheets)'
paašk-	'broken (of three-dimensional objects)'

**6. Implications for historical work.** Given the analyses and understanding of pronoun/pronominal systems that we have developed in this paper, let us examine the implications of our discoveries for the long-range comparativist. One important consequence of the facts presented in this paper should be to raise a caveat regarding the use of pronouns and pronominal affixes in long-range comparison. This paper calls into question the assumption that parallels in personal pronouns must be due to genetic inheritance if borrowing and chance resemblance are ruled out. We must also consider the pragmatic effect of backgrounding as suggested here.<sup>11</sup>

In closing let me look at a recent change in a familiar pronoun system to underline the point I have been making about pronoun systems changing in ways that bring them more in line with either the distinctness or coherence. The English pronoun system underwent a change from Early Modern English cited in 29a. to Modern English cited in 29b. The change involves the replacement of some of the second person forms by others and demonstrates one of our points—that systems prefer phonological distinctness. So while the ultimate cause of the restructuring of the system was driven by the pragmatics of politeness “inflation” with forms of the polite *ye*, replacing the familiar *thou*, I would argue that the choice of the accusative *you* in contexts requiring nominative over the older nominative *ye* was driven by the fact that *you* generates a more optimally distinct system phonologically, replacing the two-way distinction between speech act participant and third person, *a* vs. *i/i* (singular) and *i* vs. *e* (plural), with a three-way distinction in vocalism.

## 29. a. Early Modern English

old nominative	sg	pl
first person	/ay/	/wi/
second person	/ðaw/	/yi/
third person	{ /hi/ /ʃi/ /ɪt/	/ðe/

<sup>11</sup>Morphological reshaping is, of course, as widespread in this core vocabulary as anywhere else in the vocabulary.

## 29. b. Modern English

new nominative	sg	pl
first person	/ay/	/wi/
second person	/yu/	/yu/
third person	{ /hi/ /ʃi/ /ɪt/	/ɜe/

## References

Campbell, Lyle

- 1993 Problems with pronouns in proposals of remote relationships among Native American languages. Paper read to the Society for the Study of Indigenous Languages of the Americas, Columbus, Ohio, July 1993.

Forchheimer, Paul

- 1951 *The category of person in language*. Berlin: Walter de Gruyter.

Gordon, Matthew J.

- 1995 The phonological composition of personal pronouns: implications for the genetic hypotheses. *Berkeley Linguistic Society* 21:117-128.

Helmbrecht, Johannes

- 1996 On the grammaticalization of 1st and 2nd person pronominal affixes in North American Indian languages. Presented to the Special Session on American Indian Languages at the 22nd annual meeting of the Berkeley Linguistic Society, Feb. 16, 1996.

Nathan, Geoffery

- 1996 Steps towards a cognitive phonology. In B. Hurch and R. Rhodes, *Natural Phonology: the State of the Art*. Berlin: Mouton de Gruyter. 107-120.

Nichols, Johanna

- 1986 Head-marking and dependent-marking languages. *Language* 62:56-119.
- 1996 The geography of language origins. Presented to the 22nd annual meeting of the Berkeley Linguistic Society, Feb. 19, 1996.

\_\_\_\_\_ and David Peterson

- 1996 The Amerind personal pronouns. *Language* 72:336-371.

Rhodes, Richard A.

- 1996 Sound symbolism in the history of an Algonquian semantic domain. Presented to the Special Session on American Indian Languages at the 22nd annual meeting of the Berkeley Linguistic Society, Feb. 16, 1996.

Shevoroshkin, V. (ed.)

- 1990 *Proto-Languages and Proto-Cultures : materials from the First International Interdisciplinary Symposium on Language and Prehistory, Ann Arbor, 8-12 November, 1988*. Bochum : Brockmeyer.

Sweetser, Eve

- 1996 Working towards a unified understanding of linguistic change. Paper presented to the UC San Diego-UC Berkeley Cognitive Linguistics Workshop. Berkeley, CA, Jan. 26, 1996.

Wiesemann, Ursula , ed.

- 1986 *Pronominal systems*. Tübingen : G. Narr.

# Proto-Amerind \*KAPA 'Finger, Hand' and Its Origin in the Old World

Merritt Ruhlen  
Palo Alto, California

I have recently provided 58 comparisons—involving both lexical and grammatical formatives—between the Amerind family, on the one hand, and the Nostratic, or Eurasiatic, family on the other, which suggest that the Amerind family is genetically closest to Eurasiatic/Nostratic (Ruhlen 1994). In the present paper I would like to add one more lexical comparison to this body of evidence.

Let us begin with the Amerind evidence. One of the etymologies that Greenberg offered in support of the Equatorial branch of Amerind was a word for 'hand' whose characteristic shape was KAPI, or some phonetically similar form (Greenberg 1987: 88). Greenberg cited forms from the Arawa, Chapacura, Guahibo, Guamo, Katembri, Maipuran, Otomaco, and Piaroa subgroups and noted that it is the general term in the large Maipuran family. In a recent detailed comparative study of the Maipuran family, David Payne reconstructs Proto-Maipuran \**kʰapi* 'hand,' with reflexes such as Curripaco *-kapi*, Waura *-kapi*, Lokono *-kʰabo*, Resigaro *-kapʰi*, Cabiari *-kaapi*, Piapoco *-káapi*, Tariano *-kapi*, and Parecis *kahi* (Payne 1991:407). Many of these forms occur with the Proto-Maipuran first-person pronoun \**nu*- (itself the reflex of Proto-Amerind \**na*- 'I, my'), for example, Curripaco *no-kapi* 'my hand' and Cabiari *nu-kaapi* 'my hand.'

Within the larger Arawakan family, of which Maipuran is one subgroup, Greenberg cites Culino *dʒepi* 'hand' in the Arawa subgroup; Itene *kapi* 'ring finger' in the Chapacura subgroup; and San José *očepe* 'arm' in the Guamo subgroup. Finally, from other branches of Equatorial that are taxonomically equivalent to Arawakan, Greenberg adds Piaroa *čifo* 'arm' in the Piaroa branch; Otomaco *gibi* 'hand' in the Otomaco branch; Katembri *kifi* 'hand' in the Katembri branch; and Cuiva *kobe* 'arm' in the Guahibo branch. Randall Huber and Robert Reed (1992:22–23) provide fuller data on the Guahibo family that shows the root in question is well attested in the family as a whole: Playero



*pe-kóbe* 'hand' (cf. also *pe-kóbe-ši-bo* 'finger'), Guahíbo *pe-kóbe*: 'hand,' Cuíva *pe-kóbe* 'hand,' and Jitnu *pe-ko* 'hand.'

In the other South American branches of Amerind, however, this root appears to be either weakly attested, or altogether absent. Isolated examples include, in Chibchan, Manare *ukaba* 'middle finger,' Sabanero *kobaragda* 'finger,' Aruaco *abata-kabo* '5' (literally, 'one-hand'), and Sanema *polakabi* '2' (presumably to be analyzed as *pola-kabi* '2 fingers'). The only possible Paezan example I have found is Mutilon *koba* '5' (presumably, 'hand' = '5 fingers'). In the Andean branch it seems likely that Aymara *kupi* 'right hand,' Yahgan *kupaspa* '5' and Qawashqar *kuupačpe* '5' are related (note the similarity of these forms with the Playero word for 'finger' cited above). The only possible Macro-Tucanoan form I have found is Querari *yimō-kopel* 'hand.' Though Greenberg cited no Tupi forms in his Equatorial etymology—no doubt because this root is clearly not common in Tupi—there is one possible example, Aweti *ikōva* 'arm.' In Macro-Carib the lone example I have found is Maiongom *kepa'ala* 'arm'; in Macro-Panoan, possible cognates are Sanapana *ihwapesi* 'finger' (note the similarity of this form with the Playero word for 'finger' cited above) and Komlek *kovaiyi* '5.' I have found no examples in Macro-Ge languages.

In North America the situation is even more skewed—with respect to the distribution of this root—than in South America. I have found no examples at all in Almosan, Keresiouan, Hokan, or Central Amerind, and yet the root is as abundantly attested in the Penutian branch in North America as it is in the Equatorial branch in South America. We may begin our discussion of the Penutian evidence with the Mayan family (a constituent of the Mexican branch of Penutian), where the root in question is the general word for 'hand': Yucatec *kab*, Huastec *k'ubak*, Chol *k'ab*, Tzeltal *k'ab*, Tzotzil *k'abal*, Mam *kop*, Quiché *gab* ~ *q'ab*, Kakchiquel *qa*, and Pokonchi *k'ab*. For the Quichéan branch of Mayan, Lyle Campbell (1977) has reconstructed Proto-Quichéan *\*q'ab* 'hand.' In some Mayan languages, however, the meaning is 'finger': Chontal *k'ōb*, Aguacatec *vi-k'ab*, Ixil *k'ab*, and Uspantec *ba-k'ab*. Evidence from other branches of Penutian, located for the most part in the American northwest, include Tsimshian *á:pχan* 'to paw, rake, scratch,' San Juan Bautista *kupis* 'pinky,' Wintun *k'op* 'hold tight in the hand or claw; grab; claw,' Patwin *k'upum* 'finger' (apparently borrowed into Lake Miwok as *k'úpum*), Proto-Yokuts *\*xap<sup>h</sup>(a)p<sup>h</sup>al* 'finger,' Yokuts *xaphal*

'finger,' Yaudanchi *xapad* 'finger,' Nisinan *k'a:pe* 'own,' and perhaps Chitimacha (a member of the Gulf subgroup located in the Southeastern United States) *waši-ʔape* 'finger' (the first element of this compound means 'hand').

The strong evidence for this particular root in both Equatorial and Penutian, coupled with isolated instances in other branches, suggests that this root already existed in Proto-Amerind. But is it an Amerind innovation, or rather an inheritance from an even more remote time? Only the external context—the Old World—can resolve this issue, but that context does appear to supply a resolution in that the Amerind root discussed above bears a more than passing resemblance to one of Illich-Svitych's Nostratic etymologies: No. 190, Proto-Nostratic \**k'aba/k'ap*'a 'seize' (Illich-Svitych 1971: 313–15).

Illich-Svitych suggested evidence from all six Nostratic sub-families: Afro-Asiatic, Kartvelian, Indo-European, Uralic, Dravidian, and Altaic. In general, the meaning of the root is 'seize, hold in the hand'; in Kartvelian, and, in part, in Afro-Asiatic and Altaic, the meaning has apparently shifted to 'hold with the teeth, bite.' Or perhaps we are dealing with two phonetically similar, but historically distinct, roots. Illich-Svitych pointed out two phonetically similar roots in Dravidian, \**kavu-/kapp-* 'seize (with the mouth), grab' and \**kava-* 'seize (with the hand).' The former gives Tamil *kavvu* 'seize with the mouth,' while the latter gives Tamil *kavar* 'seize.'

For the Afro-Asiatic family Illich-Svitych reconstructed Proto-Afro-Asiatic \**qb-* 'seize, take, bite,' with reflexes in various branches of the family such as Arabic *qbw* 'take with the fingers' and Hebrew *qbl* 'take' in the Semitic branch; Shilha *gbi* 'bite' in Berber; Galla *qab* 'seize, hold,' Somali *qab* 'take, hold, have,' and Bilin *gab* 'hold' in Cushitic; Boleva *n-gob-u* 'I catch (a fish)' and Angas *gāp* 'tongs for taking food from the fire' in Chadic. In Kartvelian there are two morphologically related forms, a verb and a nominalization, as in Georgian *k'b-en-* 'bite' and *k'b-il* 'tooth.'

For Indo-European Illich-Svitych discussed two roots, Proto-Indo-European \**ghabh-* 'give, receive' and \**kap-* 'grasp.' The first root is responsible for forms such as Sanskrit *gābh-astis* 'hand,' Latin *habēō* 'I have,' English 'give,' and Polish *gabać* 'seize.' The second root has reflexes such as Latin *capiō* 'I take,' Greek *κάπτω* 'I seize,' Albanian *kap* 'I seize,' and English 'have.'

Illich-Svitych's Uralic reconstruction, \**kapp*Λ- 'seize,' leads

to such modern forms as Mansi *kápa*- 'seize,' Estonian *käega kaapa*- 'seize with the hand,' and Mari *kopa* 'palm, paw.' In a more recent study of Proto-Uralic, Károly Rédei (1986-88: 651-52) reconstructs two related roots for Proto-Finno-Ugric, *\*kappə* 'seize, take, grasp,' leading to Finnish *kaappaus* 'capture' and Mordvin *kapode*- 'grab quickly,' and *\*käppä* 'hand, paw,' leading to Finnish *käpälä* 'hand, paw,' Estonian *käpp* 'claw, paw, hand,' and Mordvin *kepe* 'barefooted.'

For Altaic, Illich-Svitych reconstructed Proto-Altaic *\*k'apa-/k'aba* 'seize' to account for such forms as Proto-Turkic *\*k'ap(a)*- 'seize,' Yukut *xap*- 'seize,' Azerbaijani *gap*- 'seize'; Written Mongolian *qaba/qabu* 'seize,' Khalkha *xaw* 'dexterity,' Buryat *xabatai* 'adroit, dexterous,' Kalmyk *xawɭ* 'grab, catch'; and in the Tungus subgroup, Oroch *xapki*- 'grab by the throat, strangle' and Evenki *apki*- 'strangle.' In his recent reconstruction of Proto-Altaic, Sergei Starostin (1991: 289) posits *\*k'ap<sup>h</sup>V* 'seize, hold,' and includes Proto-Japanese *\*káp*- 'buy' and its reflexes Old Japanese *kap*- 'buy' and Tokyo dialect *kà-u* 'buy.'

There is, in addition, a second Nostratic etymology that is both phonetically and semantically similar to the one just discussed, Proto-Nostratic *\*K'äp<sup>h</sup>ä* 'paw' (No. 222). Illich-Svitych provides evidence from the Afro-Asiatic, Indo-European, and Uralic families. Afro-Asiatic forms include Proto-Semitic *\*kap(p)* 'palm,' Somali *qəb* 'hoof,' Proto-Chadic *\*k'ap*- 'foot, sole, hoof,' Hausa *k'áfä* 'foot, sole' and Logone *kābē* 'hoof.' Indo-European forms derive from Proto-Indo-European *\*kap(h)o* 'hoof,' which gives Sanskrit *śaphā*- 'hoof,' Avestan *safa*- 'hoof of a horse,' Old Icelandic *höfr* 'hoof,' English 'hoof.' For Uralic, Illich-Svitych reconstructed Proto-Uralic *\*käppä* 'paw,' with the same supporting forms as those given by Rédei.

In their monograph on Nostratic, Allan Bomhard and John Kerns (1994: 404-05) reconstruct Proto-Nostratic *\*k<sup>h</sup>ap<sup>h</sup>*- 'take, seize; hand' and the supporting forms they cite overlap to a great extent with those of Illich-Svitych. Their etymology includes Proto-Indo-European *\*kap*- 'seize, take' (but not Proto-Indo-European *\*ghabh*- 'give, receive'); for Uralic, the Proto-Finno-Ugric forms from Rédei (1986-88) cited above; and for Altaic, forms similar to those cited by Illich-Svitych.

For Afro-Asiatic, Bomhard and Kerns include forms overlapping with those of the second Nostratic etymology. They reconstruct Proto-Afro-Asiatic *\*k<sup>h</sup>ap<sup>h</sup>*- 'take, seize; hand,' with reflexes such as Hebrew *kap* 'palm,' Arabic *kaff* 'palm of the

hand,' Syriac *kappā* 'palm of the hand,' and Akkadian *kappu* 'hand,' in the Semitic branch; Ancient Egyptian *kp* 'seize; hollow of the hand'; and in the Cushitic branch, Proto-Southern Cushitic *\*kip-* 'handle,' Iraqw *kipay* 'handle,' and Ma'a *-kupurúya* 'snatch.'

For Dravidian, Bomhard and Kerns suggest a different etymology than that used by Illich-Svitych, No. 1225 in the most recent edition of Burrow and Emeneau's Dravidian etymological dictionary (Burrow and Emeneau 1984: 114). This etymology is restricted to the Kurux-Malto subgroup of Dravidian—the most divergent Dravidian subgroup after the isolated language Brahui, which is both geographically and genetically the most divergent language in the family. In Kurux we find *kappnā* 'cover or press gently with the hand' and in Malto, *kape* 'touch.'

Bomhard and Kerns place the Afro-Asiatic, Kartvelian, and Dravidian forms cited by Illich-Svitych in a different etymology, No. 288, Proto-Nostratic *\*k'ab-* 'seize, bite.' All of the forms cited in this etymology, however, involve 'biting with the mouth' in some fashion, except those in Afro-Asiatic, which are exclusively 'holding in the hand,' for example, Arabic *ḡabaḡa* 'seize, hold, grasp,' Proto-East-Cushitic *\*k'ab* 'seize, take hold of,' and Proto-Southern Cushitic *\*k'ab* 'restrain.' These Afro-Asiatic forms seem to have been placed with the other—semantically quite different—forms on phonological grounds, but I would prefer to include them with the other etymology dealing with 'holding.'

Without attempting to sort out the precise relationships among all of the forms listed by Illich-Svitych, Bomhard, and Kerns, it seems clear that there is an abundance of forms in Eurasian language families—and in Afro-Asiatic—that are strikingly similar to the Amerind forms with which we began our investigation. It would thus appear that the root *\*KAPA* 'hand; seize, take' is yet one more trait connecting the Amerind family with the Nostratic/Eurasiatic family of the Old World.

## References

- Bomhard, Allan R., and John C. Kerns  
 1994 *The Nostratic Macrofamily: A Study in Distant Linguistic Relationship*. Berlin: Mouton de Gruyter.

- Burrow, Thomas, and Murray B. Emeneau  
1984 *Dravidian Etymological Dictionary*. Oxford: Oxford University Press.
- Campbell, Lyle  
1977 *Quichean Linguistic Prehistory*. Berkeley: University of California Publications in Linguistics 81.
- Greenberg, Joseph H.  
1987 *Language in the Americas*. Stanford: Stanford University Press.
- Huber, Randall Q., and Robert B. Reed, eds.  
1992 *Vocabulario Comparativo: Palabras Selectas de Lenguas Indígenas de Colombia*. Santafé de Bogotá: Instituto Lingüístico de Verano.
- Illich-Svitych, Vladislav M.  
1971 *Opyt sravnenija nostratičeskix jazykov*, Vol. 1. Moscow: Nauka.
- Payne, David L.  
1991 "A Classification of Maipuran (Arawakan) Languages Based on Shared Lexical Retentions." In: *Handbook of Amazonian Languages*, Vol. 3, eds. Desmond C. Derbyshire & Geoffrey K. Pullum. Pp. 355–499. Berlin: Mouton de Gruyter.
- Rédei, Károly, ed.  
1986–88 *Uralisches etymologisches Wörterbuch*, 3 Vols. Wiesbaden: Otto Harrassowitz.
- Ruhlen, Merritt  
1994 "The Linguistic Origins of Native Americans." In: *On the Origin of Languages: Studies in Linguistic Taxonomy*. Pp. 207–41. Stanford: Stanford University Press.
- Starostin, Sergei A.  
1991 *Altajskaja problema i proisxoždenije japonskogo jazyka*. Moscow: Nauka.

## On the "consonant splits" in Japanese

S. A. Starostin

*Russian State University of the Humanities*

Since 1992 a team of researchers in Moscow (including A. V. Dybo, O. A. Mudrak, I. N. Shervashidze and me) has been working on compiling a comparative dictionary of Altaic languages. The number of known etymologies has grown immensely, and even before the dictionary is published (which I hope will be already soon) it has become possible to considerably refine our knowledge of comparative Altaic phonology. The reconstruction of consonants has not changed much in comparison with my book of 1991, but vocalic correspondences have been significantly improved. It has also become possible to solve several phonological problems within individual branches of Altaic, and in this paper I shall try to demonstrate it for Japanese.

In [Starostin 1991, 82] a system of phonetic correspondences connecting Japanese and other Altaic languages was presented. Proto-Japanese had a rather simplified consonant system in comparison with other subgroups of Altaic, so in most cases mergers of several phonemes had occurred: thus,  $*p'$  and  $*p$  yield PJ  $*p$ ,  $*k'$  and  $*k$  yield PJ  $*k$  etc. Most Proto-Altaic consonants give simple and unambiguous reflexes in Proto-Japanese: these are the cases of PA  $*p'$ ,  $*p > PJ *p$ ; PA  $*m > PJ m$ ; PA  $*w > PJ -w-$  (sometimes weakened to  $-0-$ ); PA  $*t' > PJ *t$ ; PA  $*n > PJ *n$ ; PA  $*l > PJ *n-$  (in initial position),  $-r-$  (in medial position); PA  $*-l' > PJ *-s-$  (a rule established originally by R. A. Miller, see [Miller 1971, 114]); PA  $*č > PJ *t$ ; PA  $*č > PJ *t-$  (initially),  $*-s-$  (medially); PA  $*ž > PJ *d-$  (initially),  $*-j-$  (medially); PA  $*j > PJ -j-$  (sometimes weakened to  $-0-$ ); PA  $*k', *k > PJ *k$ ; PA  $*g > PJ *k-$  (in initial position),  $-0-$  (in medial position); PA  $*s, *z > PJ s$  (the development  $*z > s$  is not mentioned in Starostin 1991, although it appears quite regular). Another phoneme recently reconstructed (primarily on the basis of PTM  $*š$ ) is PA  $*š$  which also quite uniformly yields PJ  $*s$ .

Several PA phonemes, however, have split reflexes in Japanese. The following are the riddles of Japanese historical phonology:

- 1) PA  $*b$  yields either PJ  $*b$  ( $-w-$  in medial position) or  $*p$ ;
- 2) PA  $*t, *d$  yield either PJ  $*d$  ( $-j-$  in medial position) or  $*t$ ;

- 3) PA \*r, \*r' yield either PJ \*-r- or PJ \*-t-;
- 4) PA \*ń in initial position yields either PJ \*n- or PJ \*m-;
- 5) PA \*ŋ in medial position yields either PJ \*-n- or PJ \*-m-;
- 6) PA resonants (\*m, \*n, \*r, \*l, \*l', \*ŋ), besides standard reflexes, are sometimes dropped in medial position.

In the present paper I shall attempt to present solutions for at least some of these problems. I shall present the etymologies in a somewhat shortened manner (just the reconstructed forms in subbranches of Altaic), hoping that the Altaic dictionary containing detailed etymologies will be soon available.

### 1. Reflexes of Proto-Altaic \*b in Proto-Japanese

The following two simple rules account for the split of \*b in Japanese:

a) PA initial \*b- is preserved as \*b- before the PJ vowels -a- and -ə- (independently of their source in PA), but yields \*pin all other cases (except when there is a \*-j- in the next syllable).

b) PA \*-b- yields PJ -p- in the vast majority of cases; the medial reflex \*-w- (or -j-) is observed only after the diphthongs \*-iu-, \*-io- and \*-ia-.

Note that there seems to be an intimate connection between the preservation of voice and the preceding / following palatal glide. It is probable that original \*b was phonetically palatalized in some positions, and this palatalized \*b failed to undergo devoicing.

Consider the following examples:

A. PA \*b- > Jap. \*b- before Jap. \*-a-, -ə-

\*bĭ / \*be- 1-t p. pron. (Turk. \*bā-n; Mong. \*bi, \*min-; \*ba, \*man-; Tung. \*bi; \*bue, \*mū-n-; Kor. \*ū-rĭ) > Jap. \*bā-

\*bĕje man; self, body (Mong. \*beje, Tung. \*beje) > Jap. \*bĕ

\*bā to bind, string (Turk. \*bā; Tung. \*ba-; Kor. \*pa) > Jap. \*bā

\*bōlV to be (Turk. \*bōl-; Mong. \*bol-) > Jap. \*bār-

\*baga ( ~ -e, -ü-) wheel Kor. \*pāhōi Jap. \*ba

\*bora to divide (Tung. \*borĭ-; Kor. \*pĕrĭ-) > Jap. \*bār-

\*bialk'o to soak, gush forth (Mong. \*bulka-; Tung. \*bilkü-) > Jap.

\*bák-

\*boda body; intestines, belly (Turk. \*bod; Mong. \*boda) > Jap. \*bàtá

\*bāla child, young (Turk. \*bāla; Tung. \*bala-) > Jap. \*bàràpa(i) [Tone is irregular]

\*buka small, young; bear a child (Turk. \*bogař; Mong. \*baga; Tung. \*bogī-) > Jap. \*bàkà-

\*biūk'a side (of body), thigh (Turk. \*bikin; Mong. \*bokaur; Tung. \*bōkan) > Jap. \*bàkì

\*bak'a to watch, search (Turk. \*bak-; Mong. \*baka-; Tung. \*baka-) > Jap. \*bàkàr-

\*bok'a chain, rim (Turk. \*bukagu; Mong. \*bugu-; Tung. \*boki-) > Jap. \*baku

\*bula bad, harm (Mong. \*bala-; Tung. \*bolga-) > Jap. \*bàrə-

\*biuka rock, hill (PTM \*buga, \*buge-nse; Kor. \*pàhói) > Jap. \*bàkà

\*bido to jump, trot (Turk. \*bidi-, Mong. \*büdüri-; Kor. \*ptüi-) > Jap. \*bò(n)tér-

\*buío to pity, be sad (Tung. \*buli-, Kor. \*pur-) > Jap. \*bàsì-

\*boío ( ~ -u-) learn, be attentive (Turk. \*bolgu-; Mong. \*bolgu-ya-) > Jap. \*bàsì-p-

\*bioga place (Mong. \*baji-; Tung. \*buga; Kor. \*pá) > Jap. \*ba (OJp. ba: one of the very few words where a voiced stop was preserved in OJp., probably because of syntactics: the usual position of \*ba 'place' is after genitive \*-n(ə)-ba "place of")

Exceptions (PA \*b- > Jap. \*p- in the same position) are very few:

\*boře perish (Turk. \*buř- / \*boř-; Mong. \*bür-il-; Tung. \*bu(r)-) > Jap. \*pàrə-(m)p-

\*baža early (Turk. \*baja; Tung. \*baži-) > Jap. \*pàjáj-

\*belo pale (Mong. \*balai; Tung. \*beli; Kor. \*pàrk- (verbal root, therefore low tone)) > Jap. \*pàrà-

\*bute itch, scab (Turk. \*büt-; Mong. \*bodu(ya); Tung. \*butu-) > Jap. \*pàtákaì

B. PA \*b- > Jap. \*p- before Jap. \*-i-, \*-u-

\*bogdu paint, variegated; spot (Turk. \*bodu-; Mong. \*budu-; Tung. \*bugdi) > Jap. \*puti



\*buli to stir, shake, smear (Turk. \*bulga-; Mong. \*būli-; Tung. \*bul-) > Jap. \*pūr-

\*bōrk'i to cover, cover (Turk. \*bōrk; Mong. \*bürkū-; Tung. \*bogda) > Jap. \*pūk-

\*biuri one (Turk. \*bir; Mong. \*büri; Kor. \*piri-) > Jap. \*pitə

\*bēgi to be cold, freeze (Mong. \*beγe-re-; Tung. \*begi-) > Jap. \*pija-

\*bedu thick, large (Turk. \*bedü-k; Mong. \*bedüyün; Tung. \*burgu-; Kor. \*piri- (with secondary low tone in a verbal stem)) > Jap. \*putuā-

\*biōre give; take, collect (Turk. \*bēr-; Tung. \*bū-) > Jap. \*piri-p-

\*bāri right, straight, direct (Turk. \*ber-; Mong. \*barayun; Tung. \*bāru; Kor. \*pārā- (with secondary low tone in a verbal stem)) > Jap. \*pitā

\*bādi face, colour (Turk. \*badram "feast"; Mong. \*bad, \*badara-; Tung. \*bāda) > Jap. \*pitāpi ( / \*pitāpi) 'forehead'

\*biar[i] calf, lamb (Turk. \*buragu; Mong. \*birayu; Tung. \*biaru) > Jap. \*pitū-nsi

\*bat'i dirt (Turk. \*bat; Mong. \*bat-ga; Tung. \*batu-n; Kor. \*ptāi) > Jap. \*pi(n)ti

\*basi payment, loan (Turk. \*basig; Tung. \*basa-; Kor. \*pskūi-, pskū-) > Jap. \*pisák-

\*bar[i] wide, thick (Turk. \*barık; Mong. \*bar-; Tung. \*baru-n; Kor. \*pār) > Jap. \*pirə-

\*bius[i] to hide (Turk. \*bus-; Tung. \*busi-; Kor. \*pski-) > Jap. \*pisə-ka

\*biud[u] down, feather, curly (Turk. \*bidik "moustache"; Mong. \*buži- / \*boži-; Tung. \*bodu-ruka) > Jap. \*pi-n-kai "beard"

\*bedu platform, lid (Turk. \*böd; Tung. \*bedu-; Kor. \*ptāi) > Jap. \*puta

\*boli a k. of cedar, pine (Turk. \*böl; Tung. \*bolgikta) > Jap. \*pusi "shrubs used as firewood"

\*biole lump, knot (Turk. \*bel-ke; Mong. \*bulu; Tung. \*bul-) > Jap. \*pusi

\*bek'u a k. of fish (Turk. \*bekre; Mong. \*bekir; Tung. \*beke) > Jap. \*pu(n)ku

C. PA \*b- > Jap. \*b- before \*-j-

\*biōji to esteem (Turk. \*bāj; Mong. \*bej-le; Tung. \*buje-; Kor.

\*pài-hó-) > Jap. \*bija

\*biju be, sit (Mong. \*büj; Tung. \*bi-) > Jap. \*bú(i)-

\*beji an ungulate animal (Turk. \*biye; Tung. \*bejū-) > Jap. \*bí

D. PA \*-b- > Jap. \*-p-

\*kēbu chew (Turk. \*gēb-; Mong. \*kebi-; Tung. \*kebe ( < \*kēpu with secondary devoicing)) > Jap. \*kúp-

\*nebi new; younger relative (Turk. \*jub-ga; Mong. \*niyu-n; Tung. \*newi) > Jap. \*nípí-

\*p'abVrV swim, (over)flow (Tung. \*pawri-) > Jap. \*pápúr-

\*nábo front, in front (Mong. \*žöb < \*žeb- "straight, right"; Tung. \*nāw-) > Jap. \*māpiā

\*žībi (~e) house (Mong. \*žuw-ka; Tung. \*žūw; Kor. \*čip) > Jap. \*(d)ipiā. Falling tone in Jap. may be due to a contamination with \*ipua 'hut' (see below).

\*ebi door, yard (Turk. \*eb; Mong. \*eýde; Tung. \*ew-le; Kor. \*ip) > Jap. \*ipua

\*sābo mark, notch (Turk. \*sāb-; Mong. \*seji-le; Tung. \*saw(i)-) > Jap. \*sa(m)pak-

\*k'ābu to swell, form blisters (Turk. \*Kāp- (with assimilatory devoicing); Mong. \*kabu-; Tung. \*xawu-l-; Kor. \*kópó-m-) > Jap. \*k[ua]pu (vocalism not quite clear)

\*siba clay, to smear (Turk. \*siba-; Mong. \*siba-; Tung. \*siwa-; Kor. \*spi-ri-) > Jap. \*sápá

\*šabo to grip (with claws) (Mong. \*siýüre-; Tung. \*šawa-) > Jap. \*sápár-

\*ābi to enjoy, rest (Turk. \*abi-; Mong. \*abu-ra-; Tung. \*āw-; Kor. \*ipā-tí (with secondary low tone)) > Jap. \*ipā-p-

\*k'aba to buy, pay back (Turk. \*kabi-n; Tung. \*xaw-; Kor. \*kaphí-) > Jap. \*káp-

\*ebi (~a-, \*ibe) grain (Turk. \*ebin; Mong. \*ebe-sü "grass"; Kor. \*pjó) > Jap. \*ipí

\*ebe to carry on the back (Tung. \*ewe-; Kor. \*əp-; Jap. \*əp-)

\*sēbe to love, have fun (Turk. \*sēb; Tung. \*sebže-; Kor. \*sipí-) > Jap. \*sə(m)pa-p-

\*ebo chase, hunt (Turk. \*āb; Mong. \*aba; Tung. \*eb-te) > Jap. \*əp-

\*k'ibe ash-tree (Turk. \*Kebrüč; Tung. \*xiwa-gda) > Jap. \*kápiāru(n)tai

\*kiba a k. of foliage tree (Turk. \*Kabak; Tung. \*kiwē) > Jap.

\*kàpài

\*k'ibu handle (Turk. \*Kiben-te; Mong. \*kiyi-; Tung. \*xīw-) > Jap.

\*kúpá

\*kabari oar (Mong. \*kajiyur; Tung. \*kawri) > Jap. \*kapiara

\*mebo to shamanize, dance (Turk. \*böğü < \*böbü- (?); Tung. \*mewu-) > Jap. \*máp-

\*k'obani armpit (Turk. \*Kojun; Tung. \*xowanī) > Jap. \*kàpíná

\*sabo gift (to a chief), service (Turk. \*sabu-(r)ga; Mong. \*sibe-; Tung. \*saw-) > Jap. \*sá(m)púrap-

\*sabi a k. of big fish (Turk. \*sebrük; Tung. \*sawu-) > Jap. \*sí(m)pí

\*sībe swamped ground, swamp vegetation (Turk. \*seb-; Mong. \*siber; Tung. \*sīwe) > Jap. \*sípà / \*sipá

E. PA \*-b- > Jap. \*-w- after \*-i-diphthongs

\*siobi fish skin, gills (Tung. \*subgu; Kor. \*spám) > Jap. \*síwá

\*siubi end (Turk. \*sib-ri; Tung. \*suwe-) > Jap. \*súwá-i

\*giūbe to smoke, roast (Turk. \*gübeç; Tung. \*güw-; Kor. \*kūb-) > Jap. \*kùjú-r- (< \*kùwí-r); The root perhaps reveals a variation \*giūpe ~ \*kiūbe.

\*p'iab[o] to mince, grind, rub (Turk. \*ob-; Tung. \*pīwē-; Kor. \*pjápái-) > Jap. \*piwa-

\*žioba (~d-) weak, bad (Turk. \*jab-; Mong. \*žoba- (> Tung. \*žoba-)) > Jap. \*duawa- (with not quite clear vocalism)

\*iobu to dig, hole (Turk. \*abi-; Mong. \*ayurkai; Tung. \*ub-gā; Kor. \*op-) > Jap. \*úwa-

\*giube to hit, pound (Turk. \*Küb-; Mong. \*göbi-; Tung. \*güw-) > Jap. \*kuwa-

\*kiabu pale (Turk. \*Kuba / \*Koba; Mong. \*kubakai; Tung. \*kiawa-) > Jap. \*kúi 'yellow'

## 2. Reflexes of Proto-Altaic \*d in Proto-Japanese

In medial position the distribution of the reflexes \*-t- and \*-j- is very similar to \*-p- and \*-w-, respectively: namely, PA \*-d- yields PJ \*-j- (-0) after the diphthongs \*-iu-, \*-io- and \*-ia-, but is devoiced to \*-t- in all other cases.

A. PA \*-d- > Jap. \*-j- after \*-i-diphthongs

\***miodu** dragon (Turk. \*badruk / \*badrak "flag or pole with a zoomorphic shape"; Tung. \*muduri; Kor. \*mìrí) > Jap. \*mui "snake" (as cyclical sign)

\***biud[u]** down, feather, curly (see above) > Jap. \*pì-n-kai "beard"

\***siada** young boy or girl (Turk. \*sada; Tung. \*sida-; Kor. \*stār) > Jap. \*sai ~ \*sia

\***niadurki** (/ \*niadirku) fist (Turk. \*judruk; Mong. \*nidurga; Tung. \*nurga) > Jap. \*ninkir- "to grab". Although vocalism in this root is somewhat hard to reconstruct, a diphthong is clearly indicated by the correspondence Mong. n- : Tung. \*ñ- : Jap. n- (see below).

B. PA \*-d- > Jap. \*-t-

\***boda** body; intestines, belly (see above) > Jap. \*batá

\***mecdo** bank; earth, place (Turk. > Kaz. betkej "(steep) bank"; Mong. \*muži; Tung. \*megdī; Kor. \*màt(h)) > Jap. \*máti

\***bedu** thick, large (see above) > Jap. \*pùtuà-

\***p'agdi** foot, foot sole (Turk. \*adak; Mong. \*(h)adag; Tung. \*pagdi(-kī) > Jap. \*pitúme

\***bādi** face, colour (see above) > Jap. \*pitápi (/ \*pitápi)

\***sido** tassel, string (Turk. \*sid-; Mong. \*sižim; Kor. \*stíi) > Jap. \*si(n)tai (~ia)

\***p'agdi** (~ p-) to moisten, dip (Tung. \*pāgda-; Kor. \*ptí- ) > Jap. \*pità-

\***bogdu** paint, variegated; spot (see above) > Jap. \*puti

\***alda** fathom (Mong. \*alda; Tung. \*alda-n; Kor. \*arǎ-m) > Jap. \*ata

\***p'ādA** (~ p-) to separate; some, other (Tung. \*pādi; Kor. \*ptǎ-n) > Jap. \*patu-

\***bedu** platform, lid (see above) > Jap. \*puta

\***k'adu-** to be worn out, destroyed (Mong. \*kad-; Tung. \*xadü-) > Jap. \*kù(n)tù-r-

\***aṇda** a k. of fragrant plant (Turk. \*aṇdur; Mong. \*ažVrgana; Tung. \*an(d)ikta) > Jap. \*antūsà

\***gedi** back, behind (Turk. \*ged; Mong. \*gede; Tung. \*gedi-muk) > Jap. \*kità

\***žabda** a k. of snake (Turc. > Turkm. juvdarxā "monster, a k. of dragon" (?); Tung. \*žabdar) > Jap. \*datua

\***kadi(rV)** strong, oppressive (Turk. \*Kadir; Mong. \*keder; Tung. \*kadara-ku) > Jap. \*kitu-

\*kīdo to attend, be respectful (Turk. \*gūd-; Tung. \*kidu- ) > Jap. \*kəiāpa- (with a secondary low tone in a -p-verb)

\*p'edi energetic (Turk. \*idi; Mong. \*hide; Tung. \*pede) > Jap. \*pi(n)tua-

\*p'adi a k. of vessel (Turk. \*edil; Tung. \*padu) > Jap. \*pitu, \*pitu-ki

The distribution of reflexes in initial position, however, is different, but also very strict. It is clearly assimilative in nature:

PA \*d- becomes \*t- in PJ when followed by an original voiceless stop / fricative, \*-r- or a nasal (\*-n-, -ñ-); it yields \*d- in all other cases. The sound reconstructed as \*-r- may well have been voiceless in PA (or in early Proto-Japanese, since one of its reflexes is -t-, see below). Whether \*-n- and -ñ- could be pronounced voicelessly, is not quite clear. In any case, the effect of \*-n- following \*d- was just the same as of other voiceless consonants, and it is worth noting that PJ almost completely lacks roots with the initial sequence \*dVn- - the only case known being OJp. jani 'tar' (where j- goes back to \*ʃ-).

C. PA \*d- > Jap. \*t- before voiceless stops and \*-r-, \*-n-, \*-ñ-

\*dēp'e wave, flap; fly (Turk. \*jelpi- (with a secondary -l-, see AP 105; Mong. \*debi-; Tung. \*dep(-si)-) > Jap. \*tə(m)p-

\*daki near; follow (Turk. \*jak-in-, -l'i, \*jagu-k; Mong. \*daga-; Tung. \*daga; Kor. \*ta(h)-) > Jap. \*tiká-

\*dīl(-č'o) year; sun, sun cycle (Turk. \*jil; Mong. \*žil; Tung. \*dilačā; Kor. \*tolč; Jap. \*təsi)

\*dalp'V flat, wide (Turk. \*jalpi; Mong. \*dalba-; Tung. \*delpi-n) > Jap. \*təpira

\*dūri face (Turk. \*jūr; Mong. \*dūri; Tung. \*duru-n) > Jap. \*tura

\*diōna flat surface, land, valley (Turk. \*jān; Mong. \*denži; Tung. \*dun-se) > Jap. \*təni (irregular tone)

\*dasa to regulate, govern (Turk. \*jasa-; Mong. \*das-; Tung. \*dasa-; Kor. \*təs-) > Jap. \*təsuka- "help"

\*daño to love, be friendly (Turk. \*jaña-lč; Tung. \*daña-la-; Kor. \*təñ-) > Jap. \*tənə-

\*dāpo to endure (Turk. \*job-; Mong. \*daya-; Tung. \*dābu-) > Jap. \*təpā-

\*depu wet, soak (Turk. \*jibi-; Mong. \*debte-; Tung. \*deb-) > Jap. \*tupa-tə

\***doraka** a k. of badger (Turk. \***jorakan**; Mong. \***dorgun**) > Jap. \***tàtāké**

I know only one exception:

\***děko** burn (Turk. \***jak**-; Tung. \***deg-že-gi**-; Kor. \***tāh-jó**-) > Jap. \***dák**- (but note that a variant \***tak**- also exists).

D. PA \***d**- > Jap. \***d**- before voiced stops and resonants

\***diogi** fish (Mong. \***žiga-su**; Tung. \***žogi**) > Jap. \*(**d**)**iwuá**

\***dioge** good, better (Turk. \***jeg**-; Kor. \***tjōh**-) > Jap. \***dō**-

\***dōrV** go, walk, approach (Turk. \***jori**-/\***jüri**-; Tung. \***dūr**-) > Jap. \***dór**- (tone is irregular)

\***diulu** warm (Turk. \***jili-g**; Mong. \***dulayan**; Tung. \***dül**-) > Jap. \***dù**

\***dāli** mane; collar (Turk. \***jēl**; Mong. \***del**, \***dalaŋ**; Tung. \***deli-n**) > Jap. \*(**d**)**iári**

\***diori** a small animal (flying squirrel, bat) (Turk. \***jar**- / \***jer**-; Mong. \***žirke**; Tung. \***žurki**-; Kor. \***tārāmí**) > Jap. \*(**d**)**itāti**

\***dali** to roast, burn (Turk. \***jal**-; Mong. \***dölü**; Tung. \***dalga**-; Kor. \***tār**-) > Jap. \*(**d**)**ir**-

\***dama** (~e-) ill, sick, bad (Turk. \***jaman**; Kor. \***tām**) > Jap. \***dām**-

\***deru** to shake, sway (Mong. \***derbe**-; Tung. \***der(gi)**-) > Jap. \***dùr**-

\***dülV** night (Tung. \***dolbo**) > Jap. \***duà**, \***duà-rû**

\***dāba**- to cross (a mountain) (Mong. \***daba**-; Tung. \***dāw**-) > Jap. \***dámá** (< **daba-n**)

### 3. Reflexes of Proto-Altaic \*t in Proto-Japanese.

Superficially the reflexes of \*t in PJ are the same as the reflexes of \*d, i. e. there is a variation of \*t and \*d. There are, however, significant differences in the reflexes of \*t and \*d:

1) Word-medial \*-t- almost never yields -j- in PJ. The only example I quoted in my book (p. 71) is PA \***kēta** 'go, walk' > PJ \***kájúap**-. This root should be probably reconstructed as \***kēda**, with a secondary assimilation in Turkic (\***kēda** > \***kēta** > PT \***gēt**-); both the Mongolian and the Korean reflexes can point either to \*-t- or to \*-d-. Another apparent exception is:

\***āta** (~o) step, walk (Turk. \***āt**-; Mong. \***ada**-) > Jap. \***ājùm**- (with

irregular tone). The Turk. and Mong. forms, however, can also be compared with Jap. \**útú-r-* "move, transfer" (< \**ātu*). A contamination of two original roots (\**ada* and \**ātu*) is thus possible in Turkic and Mongolian.

Examples of the regular development are:

\**zātu* relative by marriage (Turk. \**jāt*; Mong. \**sadu-n*; Kor. \**sót*) > Jap. \**sátuá* ("adopted parents")

\**pūto* branch (Turk. \**būta-*; Mong. \**huda*; Tung. \**pota*; Kor. \**ptérki*) > Jap. \**pōta* ("log")

\**k'eto* hard (Turk. \**kat*; Mong. \**küdür*; Tung. \*(x)*etu-*; Kor. \**küt-*) > Jap. \**kátá-*

\**mote* to ask (Mong. \**möči-*; Kor. \**mūd-*) > Jap. \**mōtō-ma-*

\**kito* a k. of fox (Mong. \**küderi*; Tung. \**kitiri*) > Jap. \**kitúnái*

\**kiata* salmon, a k. of fish (Turk. \**Katir-* ( ~*d'*); Mong. \**kadarān*; Tung. \**kiata*) > Jap. \**kátú-*

\**bute* itch, scab (see above) > Jap. \**pātákai*

2) In word-initial position there is a very clear complementary distribution of reflexes, depending on the following vowel. Before front PJ vowels \**i* and \**ə* (which probably was a front \**e*-vowel in early PJ) PJ has a voiced reflex \**d-* (or \**0-* before *i*; in this position PJ does not distinguish between \**d-* > *j-* and \**0-*); before the PJ back vowels \**u* and \**a* there is a uniform reflex \**t-*. Thus we may suppose that \**t* became phonetically \**t'* here and merged with \**d'* (< \**d*); in all other positions \**t* gives the same reflexes as the original \**t'*.

A. PA \**t-* > Jap. \**d-* (0-) before Jap. \**-i*, \**-ə*

\**tiöl'i* stone (Turk. \**diāl'*; Mong. \**čilayu*; Tung. \**žola*; Kor. \**tōrh*) > Jap. \*(d)isì

\**tēri* surface, skin; color (Turk. \**deri*; Mong. \**čiraj*; Tung. \**dēre*) > Jap. \*(d)irō

\**tō-* four (Turk. \**dört*; Mong. \**dör-ben*, \**dö-čin*; Tung. \**dügin*) > Jap. \**də-*

\**telk'i* decking, duck-boards; raft (Turk. \**Tel(k)-*; Tung. \**delkē*; Kor. \**tirkuàr*) > Jap. \*(d)iká(n)ta

\**tire* to sink, enter (Turk. \**derin* "deep"; Tung. \**žiri*; Kor. \**t'ir-*) > Jap. \*(d)ir-

## B. PA \*-t- &gt; Jap. \*t

\*tēka high; top, mountain (Turk. \*dāg; Mong. \*deg- / \*deye-; Tung. \*deg-; Kor. \*tə-, \*thjə-) > Jap. \*tāká-

\*tamV root; strength, soul (Turk. \*damir; Mong. \*daŋ-gi (<\*dam-gi)) > Jap. \*tāmà

\*tiŋu listen, consider; proclaim (Turk. \*dīŋla-/\*dīŋlā-; Mong. \*duŋul-) > Jap. \*tūN-ká-

\*tiuke to pour (Turk. \*dök-; Kor. \*tahi-) > Jap. \*túk-

\*tiolu wave, shallow place (Turk. \*dalKu-; Mong. \*dolgi-; Tung. \*dol-; Kor. \*tór) > Jap. \*tù "a ford"

\*ta(w)ko a bird of prey (Turk. \*dogan; Kor. \*tawàki) > Jap. \*tāká

\*turu crane (Turk. \*durunja; Kor. \*túrúmi) > Jap. \*túrú

\*teŋu axle, spindle (Turk. \*deŋil; Kor. \*thòn) > Jap. \*tumu

\*tuki to pound (Turk. \*düg-; Tung. \*dug-) > Jap. \*túk-

\*tam[o] to drip, soak (Turk. \*dam; Kor. \*tām-) > Jap. \*támár-

\*tiopu trade, barter (Turk. \*dabar; Mong. \*düji-; Kor. \*tò'íi) > Jap. \*tupijái

\*tār[U] to pull, hang (Turk. \*dar-t-; Mong. \*tata- (< Turc.?); Tung. \*der-de-; Kor. \*tār-) > Jap. \*túr-

\*toki mound, dam (Turc. > Chag. tögä-baš "stone plate on a grave"; Tung. \*dug[i]-; Kor. \*tuk) > Jap. \*túkà

\*tak'a follow (Turk. \*daki; Mong. \*daki-; Tung. \*daka-) > Jap. \*tā(n)kápí

\*tok'a base of a horn, callosity (Turk. \*Tok; Mong. \*duku; Tung. \*dokta-) > Jap. \*takua

\*tētu respect, care (Turk. \*Tetig; Mong. \*čida-; Tung. \*dēdu-) > Jap. \*tūtù-

\*tōri birch bark, vessel made of birch bark (Turk. \*Tōr; Tung. \*duri) > Jap. \*tútú

## 4. Proto-Altaic \*ř

In my book (p. 74) I mention that after the reflex \*-ř- > -t- PJ only has vowels -i, -u. What I failed to notice, however, is that these vowels are never present after the reflex \*-ř- > -r-. The two reflexes of PA \*-ř- are, therefore, in perfect complementary distribution: \*-ř- > -t- before PJ \*-i, \*-u (whatever their origin was), but > \*-r- before PJ \*-ə, \*-a.



## A. PA \*-r- &gt; Jap. -t- before Jap. \*-u, \*-i

\*t'owurV earth, soil, dust (Turk. \*tör; Mong. \*toγur-; Tung. \*tūrV; Kor. \*tōr-) > Jap. \*tūtī

\*giafa walk, step (Turk. \*geř-; Mong. \*gar-; Tung. \*giari-/\*gira-) > Jap. \*kātī

\*niār[a] young; spring, summer (Turk. \*jār; Mong. \*nirai; Tung. \*nar-gu-; Kor. \*njērī-m) > Jap. \*nátū

\*ōri middle, inside (Turk. \*ōr; Mong. \*örü; Tung. \*uri) > Jap. \*útī

\*tōri birch bark, vessel made of birch bark (see above) > Jap. \*tútū

\*biār[i] calf, lamb (see above) > Jap. \*pitú-nsi

\*siro to leak, ooze (Turk. \*siř-; Mong. \*sir-; Tung. \*sire; Kor. \*hīri-) > Jap. \*situ 'damp, wet'

## B. PA \*-r- &gt; Jap. \*-r-

\*sāri know; beware, feel (Turk. \*seř- (~ē-); Mong. \*seri-; Tung. \*sā-; Kor. \*sari-/\*səri-) > Jap. \*sir-

\*k'iuřu red, reddish; brown, dark (Turk. \*kiř-il; Mong. \*küre-(\*kūri-); Tung. \*xuri; Kor. \*kūri) > Jap. \*kúrá-

\*niuro to become wet, soak (Turk. \*jür-; Mong. \*nor-; Tung. \*nūr- (~i-); Kor. \*(n)īr-) > Jap. \*nura-

\*čawVřV salt; bitter, acid (Turk. \*dūr (~ -lū-); Mong. \*dabu-su; Tung. \*žujar-; Kor. \*čjēr-) > Jap. \*túrá-

\*dūri face (see above) > Jap. \*túra

\*ēra to go astray, mistake (Turk. \*ār-; Mong. \*ereγü; Tung. \*eru-; Kor. \*ərjə-b-) > Jap. \*ará-

\*mařa (~e-) far, foreign (Turk. \*ba(:)ř; Kor. \*mēr-) > Jap. \*márá

\*p'ōře top (Turk. \*ūr (/ \*ōř); Mong. \*horaj; Tung. \*poro-n) > Jap. \*pórə

\*ujguřV river, small river (Turk. \*ügüř; Mong. \*üjer; Tung. \*uwgē(r)-; Kor. \*jəhīr) > Jap. \*urə

\*k'epořV curved bone (Turk. \*kebrē; Mong. \*kabir-; Tung. \*xebti; Kor. \*kupīrən) > Jap. \*kə(m)pura (~ua-)

\*bař[i] wide, thick (see above) > Jap. \*pīrə-

\*čuri string, to string (Turk. \*diř- / \*dūr-; Mong. \*dörü; Kor. \*čūr-) > Jap. \*túrá

\*aři thorn, fang (Turk. \*ařig; Mong. \*araya; Tung. \*ar-) > Jap. \*irə

### 5. Proto-Altaic \*r

In my book (p. 73) I tried to formulate a rule, according to which PA \*r > PJ \*-r- in the vicinity of rounded vowels or \*-w-, but > -t- elsewhere. Unfortunately, this rule has very numerous exceptions and is now to be abolished. It should be mentioned that nothing similar to the distribution of the reflexes of \*-r- can be observed here: both PJ \*-r- < \*-r- and PJ \*-t- < \*-r- can freely occur in front of any PJ vowels. By now this remains almost the only unmotivated split in PJ, and a possibility should be considered of reconstructing two different phonemes in PA (\*r yielding \*r everywhere, and \*r<sub>1</sub> (possibly \*r̥), yielding \*t in Japanese, but \*r in all other subgroups).

### 6. Proto-Altaic \*ń

Recently A. V. Dybo managed to demonstrate that Mongolian has also a double reflex corresponding to PTM \*ń - namely, either PM \*n- or \*ʒ-. She also demonstrated that in cases when Mongolian has \*n- here, Japanese also has \*n-; in fact, in this row of correspondences PA \*n- is to be reconstructed, with a secondary palatalized reflex \*ń- in PTM, due to the position of \*n- before a front vowel or a rising diphthong with \*-i- as the first component. The second row of correspondences (PTM \*ń- : PM \*ʒ- : PJ \*m-) actually reflects the PA palatal \*ń-.

A. PA \*n- before front vowels > Mong., Jap. \*n-, Tung. \*ń-

\*niurfo to become wet, soak (Turk. \*jür-; Mong. \*nor-; Tung. \*ńür- (~i-); Kor. \*(n)ir-) > Jap. \*nura-

\*niār[a] young; spring, summer (Turk. \*jār; Mong. \*nirai; Tung. \*ńar-gu-; Kor. \*njəri-m) > Jap. \*nátu

\*niVmV warm; soft, mild (Turk. \*jim-lčak; Mong. \*nomu- /\*neme- / \*nima-; Tung. \*ńama / \*ńem- ) > Jap. \*nàmià

\*niuńa a k. of grass (Turk. \*jon-irčga; Mong. \*nimniya; Tung. \*ńuńV; Kor. \*nàńi) > Jap. \*nàntúna

\*nik'[u] to grind, crunch; rub (Turk. \*jik-; Mong. \*niku-; Tung. \*ń[i]ki-; Kor. \*nəhír-) > Jap. \*nə(n)kə-p-

\*neri face, resemblance (Mong. \*niyur (?); Tung. \*ner-ke; Kor. \*niri-) > Jap. \*ni-, \*nər-

B. PA \*n- > Mong. \*ɣ-, Jap. \*m-, Tung. \*n-

\*nīŋči thin, narrow; short (Turk. \*jīŋč-gä; Mong. \*ɣiɣig; Tung. \*nisi-) > Jap. \*mīnsikā-

\*nām(n)Vkt'V a k. of tree (Turk. \*jīmurt; Mong. \*ɣimuyu-su; Tung. \*nāmnikta; Kor. \*nāmok) > Jap. \*mēmīti

\*nābo front, in front (Mong. \*ɣöb < \*ɣeb- "straight, right"; Tung. \*nāw-) > Jap. \*māpiā

\*nāŋŋV South (wind), warm season (Turk. \*jāj; Mong. \*naɣir (metathesis < \*ɣani-r); Tung. \*nenēne) > Jap. \*mīnāmī Mong. has a metathesis < \*ɣanir.

\*nāl(b)a young (Turk. \*jāl; Mong. \*ɣalayu; Tung. \*nalba-) > Jap. \*masu-ra-

\*nīūŋū liquid faeces (Turk. \*jin; Mong. \*ɣungag; Tung. \*nōŋŋa; Kor. \*nu(ŋ)-) > Jap. \*umī (dissimilation < \*mūmī) "pus"

\*niark'e to pinch (hair) (Turk. \*jarkak; Mong. \*ɣirge; Tung. \*nirku-) > Jap. \*mē(n)k-

\*nāŋme hundred (Mong. \*ɣayu-n; Tung. \*namā) > Jap. \*muāmua (the rare -ua-diphthong in Jap. is probably a trace of the simplified cluster)

\*nāŋa a k. of small bird (Turk. \*jaŋa-lpaj; Mong. \*ɣana; Tung. \*nāŋa-) > Jap. \*mami-

\*nāŋe nut (Turk. \*jāŋak; Mong. \*ɣiɣag; Tung. \*nāŋu-) > Jap. \*mōmō (~ -ua-) "peach"

\*nīro a k. of big fish (Mong. \*ɣirga; Tung. \*nīru- / \*neri-) > Jap. \*mōrōkō (~ -ua)

\*nū- six (Mong. \*ɣi-rgu-ya; Tung. \*nū-ŋu-) > Jap. \*mu-

\*nūt'V plant glue (Mong. \*ɣutan; Tung. \*nūte) > Jap. \*mōtī (~ -ua-)

In a few cases the \*ni- sequence was reanalyzed in Mong. as \*ni-:

\*nīā eye (Turk. \*jā-l' "tear"; Mong. \*ni-dü; Tung. \*nīā-sa; Kor. \*nūn) > Jap. \*māiN, \*mi-

\*nīama low, level; precipice (Turk. \*jamač; Mong. \*nam; Tung. \*nīama) > Jap. \*mama

### 7. Loss of resonants in Proto-Japanese

Medial consonants are usually preserved in Japanese. However, there is quite a number of cases when word-medial \*l, \*n, \*r and \*m are lost, resulting in a vowel contraction and emergence of monosyllabic words in PJ. It is interesting to note that \*l' and \*r' are never lost. In my book (p. 76) I cited two examples of an apparent loss of \*-l', but for both a different explanation is now available.

There is no strict rule predicting whether a resonant will be preserved or lost in Japanese. One should note, however, that many of the examples involved demonstrate either a morphological structure with the "disappearing" suffix \*-i (like \*pə-i 'fire', \*nu-i 'red colour', \*mú-i 'body', \*sə-i 'back') or the PJ diphthong \*-ua (in most cases when PA had an \*-u-vowel: \*dua 'night', \*sua 'hemp', \*kua 'flower', \*kua 'basket', \*kua 'child', \*kua 'silk-worm', \*tua 'door').

The possible explanation involves morphological affixation. It is quite probable that the word final \*-ua-diphthong in PJ goes back to a common PA nominal suffix \*-gV, widely reflected in other Altaic languages. Cf. matches like PJ \*kímuà (< \*k'emi-gu-) = PTM \*xemu-g-de; PJ \*sírúà = PT \*sārig; PJ \*pùtuà = Mong. bedü-gü-n; PJ \*ka(n)tua = Mong. kutu-g; PJ \*susua = Mong. sise-ge-i etc. At least in two cases a match like this can be found in words that interest us here: PJ \*kùà 'basket' = Mong. \*kori-ya 'wattle' and PJ \*kúa 'child' = Mong. keyü-ken (probably dissimilated < \*keyü-gen) id. What actually occurred was probably a loss of final vowel before the attached suffix, e. g. \*kuŋi-ga 'child' > \*kuŋga, \*kuru-ga 'wattle, basket' > \*kur-ga, with following loss of resonant in a closed syllable and finally the regular development \*-g- > -0-.

Another circumstance that speaks in favour of this explanation is that almost without exceptions the words demonstrating resonant loss are nouns. The only verbs with loss of \*-l- (and no examples of other resonants lost in verbal stems exist) are \*kə- 'come' (< \*gele,) \*á- 'receive, obtain' < \*ala and \*sə- 'make' (= Turk. \*sal- 'put'). In these cases, however, we may deal with a quite different phenomenon, because some Turkic word forms in the paradigms of these roots also reveal loss of \*-l- - which may be therefore an archaic verbal affix attached to originally monosyllabic verb roots.

There are, certainly, still unexplained features of the Proto-Japanese phonological system, such as the split of \*-ŋ- > -n- / -m- which I am still unable to explain, and the split of \*-r- > -r- / -t-

(possibly reflecting an archaic phonological distinction lost elsewhere, see above). Nevertheless, it seems now that in general the PJ system of consonants can be very well derived from the reconstructed Proto-Altaic system. The same is true for the vocalic system which I intend to discuss in some detail in another forthcoming paper.

Altaic languages, together with Uralic, are quite precious for the reconstruction of Nostratic, because they seem to preserve very well the original root structure and vowel system. Any changes in the reconstruction of Proto-Altaic (and there had been quite a few of them recently) will certainly result in modifications and improvements of our knowledge in the Nostratic area. I am quite convinced that, after the first bold attempts, we are now approaching a new stage in long-range comparative linguistics: reevaluation of what is already achieved and moving forward on the basis of improved correspondences and enlarged evidence.

#### References

1. Miller, R. A.: *Japanese and the other Altaic Languages*, Chicago-London 1971.
2. Starostin, S. A.: *Altajskaja problema i proiskhoždenie japonskogo jazyka*, Moscow 1991.
3. Dybo, A. V., Mudrak, O. A., Shervashidze, I. N., Starostin, S. A., *An Etymological Dictionary of Altaic Languages* (forthcoming).

## SOME JAPANESE ETYMOLOGIES

Alexander Vovin

*The University of Hawai'i at Mānoa*

The goal of this article is to provide Altaic etymologies for some Japanese words. Four words are traditionally considered to be of Austronesian origin or to be of unknown origin<sup>1</sup>. For the other three words, I complement an already existing Japanese-Korean etymology with further data from other Altaic languages. The words I provide etymologies for are *ki* 'tree', *numa* 'marsh', *hiru* 'leech', *tuki* 'moon', *dare* 'who', *te* 'hand' and *mono* 'thing'.

Japanese *ki* 'tree' < Proto-Japanese (PJ) \*k<sub>Q</sub>-Ci 1.3a (Martin 1987: 449) is traditionally believed to be a word of Austronesian origin (< Proto-Austronesian (PAN) \*kaju or \*káS<sub>2</sub>iw 'tree' (Murayama 1975: 66)). However, there is a word 𣎵𣎵 'tree' in the Koguryo fragments which should be probably reconstructed as \*k<sub>ē</sub>nēt or \*k<sub>ē</sub>nēr. Thus, the Austronesian etymology for PJ \*k<sub>Q</sub>- < \*k<sub>Q</sub>ŋŋr 'tree' must be rejected. Taking into consideration this Koguryo form, Whitman's law of medial \*-r- loss in pre-PJ (Whitman 1985: 21-24) and the fact that the word in question has low pitch in PJ, which may reflect earlier vowel length (Martin 1987, 250-252), I suggest the following evolution of PJ \*k<sub>Q</sub>- 'tree' < pre-PJ \*k<sub>Q</sub>ŋŋ- < \*k<sub>Q</sub>ŋŋr- < \*k<sub>Q</sub>ŋŋŋr- < \*k<sub>Q</sub>ŋŋŋŋr- (cf. a similar development of PJ \*pari 'needle' < \*parari < \*panar-[C]i (cf. MK *panol* 'needle', where the loss of the final voiced consonant was blocked by the suffix -[C]i). The alternative etymology will be the Proto-Manchu-Tungus (PMT) \*xiñee- 'bird-cherry tree' (Tsintsius 1975a: 318). It is interesting to note that Tungusic bird-cherry tree and famous Japanese *sakura* belong to the same species. It is obvious that the word *sakura* is the euphemism, 'bloom thing' (Martin 1987: 517). Thus, it is quite possible that PJ \*k<sub>Q</sub>- < \*k<sub>Q</sub>ŋŋŋŋr 'tree' was originally a name for *sakura* and later became a designation of tree in general. I reconstruct PA \*k'iñēē kind of cherry tree'.

There is a well known MK parallel *nuph* H 'marsh' for PJ \*numa 2.3 'marsh', 'swamp' (Martin 1987: 502) proposed by S. E. Martin (Martin 1966: 236). This binary comparison may be well extended to other Altaic languages: PMT \*lewee 'swamp' (Tsintsius 1975a: 514) and PM \*lobV-qu 'swampy land' > (WM *lobqu* 'swampy land', 'very wet land which is difficult to plough' (Lessing 1995: 517), Khalkha *lovx* 'marshy or swampy land, unsuitable for agriculture' (Hangin 1986: 291). This is another example for PA initial \*l-, proposed by V. M. Illich-Svitych (Illich-Svitych 1974: XVI-XVII). I reconstruct PA \*lubä 'swamp', 'marsh'. The other examples with a correspondence of PMT \*l- to PJ \*n- will include PMT \*lamu 'sea' and PJ \*nami 2.3 'wave' (Martin 1987: 492),

<sup>1</sup>I adopt the reconstruction of three series of stops for PA (voiceless unaspirated, voiceless aspirated and voiced) as proposed by Illich-Svitych (1971: 169) and V. I. Tsintsius (1975: 299-306).

PMT \*luk- 'to take off (clothing)' (Tsintsius 1975a: 507) and PJ \*nuk-<sup>2</sup> B 'take off (clothing)'.

OJ *piru* 'leech' was compared by John Whitman with MK *pyelwok* LH 'flea' (Yu 1987: 382) < \*pilo-k (Whitman 1985: 212). I believe it is possible to add to this comparison PMT \*piru 'parasite' (> Manchu *fiyaru* 'worm', Ulchi *piru*- 'bug', moth'; Nanai *piro* 'moth', Oroch *parawu* 'leech', 'tick' (Tsintsius 1977: 37). The Tungusic word was previously compared by both Ramstedt with Korean *pelley* 'worm' (Ramstedt 1949: 174, Starostin 1991: 297), both unaware of the fact that MK *pelGey* LH 'worm' (Yu 1987: 379) includes a -G- indicated by the MK 'syllabification' (though Starostin assumes a hypothetical (Ñ) in his PK \*për(Ñ)ëi, which is, of course, unsubstantiated by internal evidence).

OJ *tukiy* 'moon' (< PJ \*tuku-Ci 2.3) is traditionally compared with MK *tol* H 'moon' (Martin 1966: 236). It is necessary to note, however, that OK has *Tolal*-<sup>3</sup> 'moon' as attested by Hyangga texts (Kim 1986: 80). Undoubtedly, the Japanese-Korean comparison is going to work only if one assumes a cluster \*-RK- in the protolanguage.<sup>4</sup> I believe that it is possible to expand further into Altaic this etymology. WM *tergel*- 'become full (of moon)' (Lessing 1995: 805) may seem at first a not very likely candidate both semantically and phonetically (back vocalism in Japanese and Korean words, but front vocalism in Mongolian), but the word *tergel* is also attested as a noun in the following compounds: *tergel sara* 'full moon' (Lessing 1995: 674), *tergel edür* 'fifteenth day of a lunar month' (Lessing 1995: 296). The last compound demonstrates that it is likely that the archetype meaning of *tergel* was '\*full moon', and therefore the semantic difference does not seem to be unpassable. The word is also attested in the *Yuan Chao Bi Shi* ("Secret History of Mongols") in a compound *hula'an tergel udur* 'der Tag "Roter Glanz"' (the sixteenth day of the fourth lunar month) (Haenisch 1939: 149). The meaning 'shining' here does not contradict the suggested etymology either.

Japanese *dare* 'who' < PJ \*ta-raCi 2.1 (Martin 1987: 391). Recently suggested Austronesian etymology is PAN \*tsa<sub>123</sub>yi 'who' (Benedict 1990: 259). I believe, however, that an alternative Altaic etymology may be proposed. In January of 1990, J. R. P. King, now of University of British Columbia, and I recorded the forms *du-gu* HL, *n<sup>h</sup>du-gu* HL 'who' from a couple of Soviet Korean informants. These forms correspond to Standard Seoul Korean *nwu-kwu* and Middle Korean (MK) *nwu* H 'who'. We believe that this initial *d-* in Soviet Korean corresponding to Seoul Standard Korean and MK *n-*, may reflect Proto-

<sup>2</sup> Martin reconstructs PJ \*nuka- (Martin 1987: 738).

<sup>3</sup> The reconstruction of the first syllable is tentative as it is written by the logogram 'moon'.

<sup>4</sup> Other examples supporting a reconstruction of this cluster: OJ *kakyi* 'oyster' (Omodaka 1967: 167) and MK *kwul* H 'id.' (Martin 1966: 238), OJ *kakey-* 'hang' and MK *kel-* R 'id.' (< PK \*kelV-) (Martin 1966: 98), OJ *puk-* 'to blow' and MK *pul-* L 'id.' (Martin 1966: 226).

Korean (PK) \*d-, since there are two sets of correspondences in Soviet Korean for Seoul and MK *n-*:

<u>gloss</u>	<u>Seoul</u>	<u>MK</u>	<u>Soviet Korean</u>
'four'	neys	neyh	døi <sup>n</sup> døi
'song'	noray	nworay	doræ <sup>n</sup> doræ
'butterfly'	napi	napwoy	dabi <sup>n</sup> dabi
'day'	nal	nal	nari
'eye'	nwun	nwun	nuni
'age'	nai	nahi	nai

We tentatively reconstruct PK \*d- on the basis of the correspondence of Soviet Korean *d-* : MK and Seoul *n-* and PK \**n-* on the basis of the correspondence of Soviet Korean *n-* : MK and Seoul *n-*. Thus, I reconstruct PK \**dwu* 'who'. A comparison of PK \**dwu* 'who' with PJ \**ta-* 'who', which I suggest here will face unsurmountable problems within comparative Korean-Japanese. This, however, may be resolved within comparative Altaic. The first obvious problem is a correspondence of PK \**d-* to PJ \**t-*. This correspondence in addition to already known correspondences PK \**t* : PJ \**t* < Proto-Japanese-Korean (PJK) \**t* and PK \**t* : PJ \**d* < PJK \**d* will suggest a new dental stop in PJK. Moreover, taking into consideration PK \**deC-i* 'four' (see the chart above) and PJ \**dq-* 'four', we come across yet another correspondence: PK \**d-* : PJ \**d*. Thus we are faced with the necessity to reconstruct for PJK four dental stops! Nevertheless, the situation is not so desperate, because both PJ \**t* and \**d* as well as both PK \**t* and \**d* may reflect PA \**t* with yet unknown distribution:

<u>PA</u>	<u>PJ</u>	<u>PK</u>	<u>PT</u>	<u>PM</u>	<u>PMT</u>
*t'	*t	*t	*t	*t	*t
*t	*t/*d	*t/*d	*d	*d	*d
*d	*d/*t	*t	*j	*d	*d

Examples: PA \**tur<sub>1</sub>u* 'crane' > PJ \**turu* 2.5 'crane', PK \**twulwu-mi* HHH 'crane', PT \**dur-na* 'crane'; PA \**tō[r]*- 'four' > PJ \**dq-* 1.1 'four', PK \**deC-i* LH 'four', PT \**dör-t* 'four', PM \**dör-* 'four', PMT \**di-* 'four'. Thus, I suppose that PK \**d-* in \**dwu* 'who' and PJ \**t-* in \**ta-* 'who' both may reflect PA \**t-*. The correspondence in vocalism may also seem strange, but there is another good example with correspondence of PJ \**a* to PK \**wu*: PA \**kat'a* 'strong', 'hard' > PJ \**kata-* 'strong', 'hard'; PK \**kwut-* L 'strong', 'hard'; PT \**kat-* 'strong', 'hard'; PMT \**kata-* 'strong', 'hard'; PM \**kata-* 'strong', 'hard'. Thus, I reconstruct PA \**ta-* 'who' > PJ \**ta-* 'who', PK \**dwu-* 'who'.

Japanese *te* 'hand' < PJ \**ta-Ci* 1.3a (Martin 1987: 545) is traditionally considered a word without an Altaic etymology. Most widespread among scholars is the acceptance of the Austronesian etymology: PAN \**taNan* 'hand' (Murayama 1974, 113-114). This etymology may be criticized phonetically: it is



unclear why Japanese lost the second syllable. Moreover, the PAN word seems to be attested only in Indonesian languages (Starostin 1991, 108). Starostin suggested that the word may be a loanword from Ainu *tak* (sic) 'hand' (Starostin 1991, 108). However, there is no word *tak* in Ainu with the meaning 'hand', the word for 'hand' in Ainu is *tek* < Proto-Ainu \*tèk (Vovin 1993: 143). In addition to the difference in vocalism, the loss of the final consonant in Japanese also represents a serious problem, provided the word *te* has the Ainu origin (we would expect a Japanese form to be something like *tekV*, cf. the adaptation of Chinese loanwords with final consonants). The hypothesis that PJ \**ta-Ci* may be a loanword from Proto-Viet-Muong \**saj* 'hand' (Starostin 1991: 108) seems to be even more fantastic.

I suspect that PJ \**ta-Ci* 1.3a 'hand' may be related to PJ \**itu-* 2.3 'five' (notice also that these two words belong to the same low register). Let us look at words in different Altaic languages with that have the meanings 'hand', 'five', 'fifty'.

	<u>OJ</u>	<u>MK</u>	<u>Ewenki</u>	<u>Manchu</u>	<u>OT</u>	<u>WM</u>
'hand'	te	swon H	Ńaale	gala	äl/älig	—
'five'	itu-	ta-sos LH	tunŃa	sunja	—	ta-bun
'fifty'	iswo-	swuyn R	tunŃa jaan	susai	älig	ta-bin

Several commentaries are necessary for this chart.

- 1) The common origin of OJ, MK, MT, OT and WM words for 'five' has already been indicated (Miller 1971: 221).
- 2) OJ *iswo-* 'fifty' is traditionally analyzed as *i-* 'five' + *swo* 'ten' (cf. OJ *myi-swo-* 'thirty', *ya-swo-* 'eighty', *i-po-* 'five hundred' etc.). However, the existence of such forms as OJ *sa-tukiy* 'fifth lunar month', where *sa-* is 'five' and *tukiy* is 'month' and absence of any other compounds where *i-* means 'five' makes me believe that OJ *iswo-* might be a truncation of an earlier *iswo-swo-* 'fifty' and *i-po-* 'five hundred' might be derived by analogy with *iswo-* reanalyzed as *i-swo-*. Thus, OJ *itu-* 'five' and *iswo-* 'fifty' are based on the same PJ root \**iTu-* 'five', where capital /T/ denotes an internal correspondence /t/:s/, unique to Japanese.
- 3) Ewenki and Manchu display the same correspondence /t/:s/, which is also unique for the Manchu-Tungus languages. However this time it occurs between different languages and not between the words derived from the same root in a single language.
- 4) One can also notice an obvious similarity between MK words for 'hand' and 'fifty'. However, since their accentuation is different, one can consider it to be sheer correspondence. But here we notice the same similarity between OT words *äl[ig]* 'hand' and *älig* 'fifty'. These words are certainly related not to MK, but to the PMT \**Ńaala* 'hand' (Starostin 1991: 17), represented in the chart above by Ewenki *Ńaale* and Manchu *gala* 'hand'. The double occurrence of the same parallelism 'hand' : 'fifty' may be explained either by a linguistic miracle or by the same productive model.

- 5) If we return to Japanese we will see that we have here the similar parallelism 'hand' : 'five' : 'fifty'. Consequently, one can suggest that MK *ta-* 'five' should be also considered as a part of the etymon.
- 6) Finally I come to the conclusion that OJ *te* 'hand', *itu-* 'five', *iswo-* 'fifty'; MK *swon* 'hand', *ta-* 'five', *swuyn* 'fifty'; Manchu-Tungus *tun-/su[n]-* 'five', and Mongolian *ta-* 'five', 'fifty' are all derived from the same PA root with a probable basic meaning 'five'. This hypothesis faces a unique 'floating' correspondence /t/:s/ not only between the different Altaic groups, but also within Japanese, Korean and Manchu-Tungus. The vocalism of the basic form also remains unclear. However, even if we remove the Japanese and Korean words meaning 'hand' from this etymon, it will not improve the situation with phonetic correspondences.
- 7) It seems that a reconstruction of a new PA phoneme in order to explain this strange correspondence will be redundant, since there is not even one single example of the 'floating' /t/:s/ correspondence between Altaic languages. I dare to venture the hypothesis that Japanese form \**itu-* 'five' with initial /i/ may be close to PA archetype. Thus, I suggest that it may be this initial /i-/ that caused a sporadic palatalization /t / > /s/ and then itself disappeared everywhere except Japanese.

In conclusion, I would like to provide an etymology for Japanese *mono* 'thing' < PJ \**mōn* 2.3 (Martin 1987: 485). To the best of my knowledge, I have not seen any etymologies of it. One obvious parallel is Early Modern Korean *mwon* 'thing' (Yu 1987: 323). The attestation is unique: this word is attested only in the "Twongen ko.lyak", a text attributed to the years of king Cengcwo's rule (1777-1800 A.D.) (Yu 1987: 813). This Korean word is not attested in MK, nor did it survive in any modern dialects. Nevertheless, the phonetic and semantic fit between PJ \**mōn* 'thing' and Early Modern Korean *mwon* 'thing' is ideal. It is not completely clear whether this PJK \**mono* 'thing' may be related to PM \**mōn* 'right', 'same', 'essence' (WM *mōn*, Khalkha *mōn*, Buriat *mōn*, Kalmyk *mōn*) and PT \**bun-* 'this' (for the comparison of Mongolian and Turkic forms see (Ramstedt 1952: 75), (Sevortian 1978: 226-227)), since in spite of a good phonetic fit, the semantics is problematic. I tentatively reconstruct PA \**mono* 'thing', 'essence'.

## ABBREVIATIONS

H	high pitch	PAN	Proto-Austronesian
L	low pitch	PJ	Proto-Japanese
MK	Middle Korean	PJK	Proto-Japanese-Korean
MT	Manchu-Tungus	PK	Proto-Korean
OJ	Old Japanese	PM	Proto-Mongolian
OK	Old Korean	PMT	Proto-Manchu-Tungus
OT	Old Turkic	PT	Proto-Turkic
PA	Proto-Altaic	WM	Written Mongolian

## REFERENCES

- Benedict, Paul K.  
1990 *Japanese/Austro-Tai*, Karoma Publishers, Ann Arbor.
- Haenisch, Erich  
1939 *Wörterbuch zu Manghol un Niuca Tobca'an* (Yuan Ch'ao Pi-Shi, Geheime Geschichte der Mongolen). Leipzig: Otto Harrassowitz.
- Hangin, Gombojab (with John R. Krueger, Paul D. Buell, William V. Rozycki, Robert G. Service)  
1986 *A Modern Mongolian-English Dictionary*. Indiana University: Research Institute for Inner Asian Studies.
- Illich-Svitych, Vladislav M.  
1971 *Opyt sravneniia nostraticheskikh iazykov* [An Attempt of Comparison of the Nostratic Languages], Nauka, Moskva.
- Kim, Wancin  
1986 *Hyangka haytok pep yenkwu* [A Study in Decipherment of the Hyangka]. Seoul: Seoul tayhakkyo chwulphanpu.
- Lessing, Ferdinand (ed.)  
1995 *Mongolian-English Dictionary*, compiled by Mattai Haltod, John Gombojab Hangin, Serge Kassatkin and Ferdinand Lessing. Third printing. Indiana University: The Mongolia Society.
- Martin, Samuel E.  
1966 "Lexical Evidence Relating Korean to Japanese", *Language* 42, #2, pp. 185-251.  
1987 *The Japanese Language Through Time*, Yale University Press, New Haven and London.
- Miller, Roy A.  
1971 *Japanese and the Other Altaic Languages*, The University of Chicago Press, Chicago and London.
- Murayama, Shichirō  
1974 *Nihongo no kenkyū hōhō* [A Method of Investigating the Japanese Language], Tokyo.  
1975 *Kokugogaku no genkai* [The Limited World of Japanese National Linguistics], Tokyo.
- Ramstedt, Gustav J.  
1949 *Studies in Korean Etymology*. Suomalais-Ugrilaisen Seuran Toimituksia 95, Suomalais-Ugrilainen Seura, Helsinki.

- 1952 *Einführung in die Altaische Sprachwissenschaft, 2, Formenlehre*, Suomalais-Ugrilaisen Seuran Toimituksia 104:2, Suomalais-Ugrilainen Seura, Helsinki.
- Sevortian, Ervand V.  
 1978 *Etimologicheskii slovar' tiurkskikh iazykov, tom 2* [An Etymological dictionary of the Turkic languages, vol 2], Nauka, Moskva.
- Starostin, Sergei A.  
 1991 *Altaiakaia problema i proiskhozhdenie iaponskogo iazyka* [The Altaic Problem and Origins of the Japanese Language], Nauka, Moskva.
- Tsintsius, Vera (ed.)  
 1975a *Sravnitel'nyi slovar' tunguso-man'chzhurskikh iazykov, tom 1* [A Comparative Dictionary of the Manchu-Tungus Languages, vol. 1], Nauka, Leningrad.  
 1977 *Sravnitel'nyi slovar' tunguso-man'chzhurskikh iazykov, tom 2* [A Comparative Dictionary of the Manchu-Tungus Languages, vol. 2], Nauka, Leningrad.
- Tsintsius, Vera I.  
 1975b "On the Pre-altaic System of Consonants", *Researches in Altaic Languages*, Akadémiai Kiadó, Budapest, pp. 299-306.
- Vovin, Alexander  
 1993 *A Reconstruction of Proto-Ainu*, E. J. Brill, Leiden.
- Whitman, John B.  
 1985 *The Phonological Basis for the Comparison of Japanese and Korean*, Harvard University Ph.D. Dissertation, Cambridge.
- Yu, Changton  
 1987 *Yico e sacen* [A Dictionary of the Korean Language During Li Dynasty], Yonsey tayhakkyo chwulphanpu, Seoul.